3-26-2020

Validation Technique for Modeled Bottomside Ionospheres via Ray Tracing

Kevin S. Burg

Validation Technique for Modeled Bottomside Ionospheres via Ray Tracing

THESIS

Kevin S. Burg, 1st Lt, USAF

AFIT-ENP-MS-20-M-085

DEPARTMENT OF THE AIR FORCE
AIR UNIVERSITY

AIR FORCE INSTITUTE OF TECHNOLOGY

Wright-Patterson Air Force Base, Ohio

AFIT-ENP-MS-20-M-085

VALIDATION TECHNIQUE FOR MODELED BOTTOMSIDE IONOSPHERES VIA RAY TRACING

THESIS

Presented to the Faculty

Department of Engineering Physics

Graduate School of Engineering and Management

Air Force Institute of Technology

Air University

Air Education and Training Command

in Partial Fulfillment of the Requirements for the

Degree of Master of Science in Applied Physics

Kevin S. Burg, B.S.

1st Lt, USAF

March 2020

AFIT-ENP-MS-20-M-085

VALIDATION TECHNIQUE
FOR
MODELED BOTTOMSIDE IONOSPHERES
VIA
RAY TRACING

THESIS

Kevin S. Burg, B.S.
1st Lt, USAF

Committee Membership:

Maj D. Emmons, Ph.D.
Chair

Dr. E. Dao
Member

Lt Col O. Nava, Ph.D.
Member

AFIT-ENP-MS-20-M-085

Abstract

A new method for validating ionosphere models using High Frequency (HF) angle of arrival (AoA) data is presented. AoA measurements from a field campaign held at White Sands Missile Range, New Mexico, USA in January 2014 provide the actual elevation angle, azimuth and group delay results from 10 transmitter-receiver circuits. Simulated AoAs are calculated by ray tracing through the electron density profiles predicted from the ionosphere models hosted by NASA's Community Coordinated Modeling Center: IRI-2016, USU-GAIM, GITM, CTIPe, TIE-GCM, and SAMI3. Through the implementation of metrics including Mean Absolute Error, Prediction Efficiency, Correlation Coefficient, and others, we are able to compare and analyze model performance using this new model validation approach. The wide variety of metrics determine model performance quantitatively and qualitatively by analyzing behavior over time.

Acknowledgements

To the most important people in my life, I wish to share my appreciation to my wife and my son. There have been long days away from you two. There have been extended TDYs and late nights working over concepts. There have been obstacles we have encountered, but there has been nothing that stopped us from achieving our goals. You two are my strength, my joy, and my loves. For that, I thank you.

This project has been a wonderful reminder of what can be accomplished when several people work towards a common goal. Though a lot of support goes to the amazing people at the CCMC (Ja Soon Shim, Lutz Rastaetter, and Katherine Gracia-Sage) and AFRL (Eugene Dao), there is no one person who deserves my thanks than Maj Dan Emmons. From the very beginning, he brought me into this multi-faceted project and guided me through foreign concepts, not only in physics, but in coding as well. Maj Emmons, you are a truly patient, caring, and brilliant individual. I have been blessed with you as an advisor and confidant.

Kevin S. Burg

Table of Contents

List of Figures

List of Tables

I. Introduction

Since the time Guglielmo Marconi realized that importance of radio waves, a focus has been made on sending data across vast distances. Now, more than a century after the first successful transatlantic radio signal, the world is connected better than ever before with people communicating to others faster than ever before. Aside for communication, the ionosphere is also vital in such applied practices as high frequency (HF) geolocation and over-the-horizon radar (OTHR), but are there models in use today that can accurately predict the state of the bottomside ionosphere?

Using a select number of models from partners at the Community Coordinated Modeling Center (CCMC) located at NASA Goddard Space Flight Center (GSFC), and a new HF ray tracer developed by partners at the Air Force Research Laboratories (AFRL) at Kirtland Air Force Base, a new technique was developed to validate these models comparing to ground-truth measurements from an experiment over White Sands Missile Range in New Mexico. This experiment, funded by Intelligence Advanced Research Projects Activity (IARPA), sent HF signals from 8 different transmitters and then picked up by a receiver. The resulting data would then be classified as Angle of Arrival (AoA) data, which is a compilation of elevation and azimuth angles, in addition to group delay. With the ground truth information and measured ionograms, we can compare the electron density profiles (EDPs) of region to the model-derived EDPs.

The EDPs show visual differences in altitudes and intensities of various features of interest in the h_mF_2 and N_mF_2 as well as h_mE and N_mE. These plots may be able to indicate some information regarding the state of the ionosphere, but they don't tell the entire story of its impacts. The ray tracer from AFRL allows for a ray to be sent through these theoretical profiles and reveal their own AoA data to then compare with the ground truth obtained during the IARPA experiment.

As a result of necessity, models were developed to assist decision makers and physicists in quantifying the ionosphere. However, a reliable model is only as trustworthy as the rigorous amount of validation the model underwent. In addition to basic validation efforts, technology has aided the advancement of understanding with regards to the ionosphere. Several works have been completed utilizing differing methods, newer technologies, and varying locations to quantify the validity of the models' ability to describe this important layer.

McNamara (1985) worked on ensuring that Total Electron Count (TEC) is a valid parameter by virtue of comparing TEC over differing latitudes. He found that some models fell short in low latitudes, but improvements could be made with numerical. Ferguson (1995) verified model validation using differing frequencies in the Very Low Frequency (VLF) and Low Frequency (LF) range of the radio spectrum. Exploring the ionosphere using lower frequencies (10 kHz-60 kHz) helped to further understand the utility of the ionosphere through different disciplines. The purposes behind this particular paper were focusing on both signal propagation and model capabilities. Anderson et al. (1998) brought us a study of the F_2 peak and determining where some physics-based models appeared to fall short. Overall, the discovery was made that the diffusion coefficient portion of a Burnside factor needed to be re-calibrated and increased. It also revealed the surprising trend that uncoupled models appeared to out-perform coupled models. Araujo-Pradere and Fuller-Rowell (2002) involved a validation effort for the F-region during storming events. This "storm-time correction model" aided with the challenging task of modeling the state of the ionosphere during its more chaotic times. The authors had a clear vision of how they wanted to display the abilities of using STORM by setting up specific criterion on specific days. The validation effort yielded favorable results regarding improving the detection of ionospheric variability due to storming.

GAIM had been a longstanding, well performing model, but there was a new addition which needed its own validation effort. Scherliess et al. (2006) performed a validation effort to ensure a Gauss Marker Kalman Filter (GMKF) was a valuable addition. This filter allowed GAIM to assimilate slant Total Electron Content (TEC) through numerous sources including GPS ground stations, electron densities from DMSP satellites, as well as UV radiance measurements. With the involved addition to the model, the validation assured users of its quality performance especially in the data-rich environment where the experiment took place. A second validation for GAIM had occurred a year later under the efforts of Decker and McNamara (2007). GAIM had been tested for its ability to perform as a GLOBAL model in Australia. As the GMKF validation showed, GAIM works quite well in a data-rich environment and this validation effort corroborated that analysis, however it was also seen that in data-sparse regions of Australia where GPS data was unavailable for assimilation, results were seen to be poorer by comparison.

Reinisch (2011) utilized more advancing technology and an ever-increasing presence in the space domain to undertake a study in the validity of a Global Ionospheric Radio Observatory (GIRO) suite of digisonde sounders whose goal was to dramatically improve accessibility to real-time ionospheric data. GIRO accomplishes this astounding feat by allowing universal access to millions of ionospheric measurement records collected at 64 locations, via over 40 real-time feeds. Shim et al. (2011) explored a large scale validation effort with Incoherent Scatter Radar (ISR) and vertical drifts quantified by several metrics. This validation effort was has several similarities to this thesis with regards to some models used and some metrics explored including Prediction Efficiency and Ratio Range. The GEM (Geospace Environment Modeling) Challenge explored different storm strengths over different latitudes with different models. Nickisch et al. (2015) brought another technological leap in a fine-tuned,

high resolution ionospheric model addressed as GPSII (Global Positioning System Ionospheric Inversion) which is a well-performing local ionosphere model. In addition to showing what a model is capable of with regards to nearly perfectly matching AoA data over a small geographic region, this paper also inspired a couple different plots such as the deviation from zenith and a Cumulative Distribution Function. Ionospheric models are not limited to testing and validation efforts in Australia and the United States alone. China's Yang et al. (2018) released a study into the testing of a Global Navigation Occultation Sounder (GNOS) which had been launched into orbit on-board a geo-satellite, FengYun 3 C. This particular validation was to quantify the accuracy of radio occulation data from both a BeiDou System (BDS) and GPS source. With up-to-date technology, the results of 0.97 correlation coefficient for F_2 peak density and 0.89 for F_2 peak altitude shows the accuracy models are capable of.

Several of the above studies had focused on model validation solely using the F_2 peak as the main parameter of focus. Fewer studies have been done with regards to confirming model performance in the bottomside ionosphere and E-region. One such study was performed by Solomon (2006), which explored comparing IRI data with values from EUVAC and Hinteregger (HFG). The comparison specifically focused on E-region ionization rates in response to soft X-rays and H Lyman-β line. Overall, the paper found that EUVAC and HFG solar models underestimated the ionization rates of the E-region, but that may be a result of relying too heavily on the F10.7 values used to drive the models. The importance of studying a model's ability to correctly identify the intensity and altitude of the F_2 peak is one that can't be overstated, but the lack of focus towards the bottomside ionosphere can lead to degradation of performance for HF applications. Transmitted frequencies interact with the ionosphere related to Snell's Law, but the refraction of said frequencies does not simply follow a mirror-like bounce when it reaches an appropriate intensity. Also seen is that the overall shape

of the ionosphere is not a flat, consistent layer for each intensity. There are areas of divots, bumps, and grooves which can greatly deflect a signal away from the originally intended path. If one were to solely fixate on F_2 data, they are capable of missing out of vital intricacies seen at lower altitudes. A new methodology is required to ensure that global models used today are capable of correctly mapping and navigating the bottomside ionosphere.

The purpose of this project is to display a new technique of validating ionospheric models and testing their efficiency. Given quiet conditions (no TIDs, consistent & available data, and no geomagnetic activity), we are giving these models their best chance to match up with the measured values. There is, at no point, a bias towards determining a best or worst model. We have simply developed a methodology which can be repeated for a model of your choosing. With this technique, we can analyze the bottomside ionosphere for a real-world application.

II. Background

In this chapter, I will be introducing a selection of ionospheric models housed at CCMC and providing some details about their inputs/outputs and how they are different from other models. I will then walk you through the Jones-Stephenson ray tracer AFRL implemented. The ray tracer used in this technique was vital for taking the information from the models and presenting it into usable data. Once there is an understanding of how we take model outputs and run them through a ray tracer, I will explain two specific examples of previous validation efforts which were instrumental in the formulation of this technique. Finally, I will describe the experiment in WSMR where the measured data came from.

2.1 Models Used

Prior to explaining the process used to verify ionospheric models, it would be beneficial to explain a few finer details regarding the models we used. This project was performed in cooperation with CCMC where they house numerous models which aid in determining the state of the magnetosphere, the heliosphere, and among others, the ionosphere. For the ionosphere, CCMC is able to run 6 models used for this project. The models used include IRI-2016 (International Reference Ionosphere), USU-GAIM (Utah State University Global Assimilation of Ionospheric Measurements), CTIPe (Coupled Thermosphere Ionosphere Plasmasphere Electrodynamics), TIE-GCM (Thermosphere Ionosphere Electrodynamics General Circulation Model), GITM (Global Ionosphere Thermosphere Model), and SAMI3 (SAMI is Another a Model of the Ionosphere, 3-dimensional). Out of the library of 12 possible ionospheric models available through CCMC, these 6 allow us to determine the electron density

profiles of a particular location over a period of time (Chulaki, 2019).

SAMI3 was developed by Joseph Huba, Glenn Joyce, and Marc Swisdak in conjunction with Naval Research Laboratory (NRL) and Icarus Research Inc. This physics-based model is a 3-dimensional global ionospheric model which uses daily and 3-month average F10.7 values, the AP Index and an E×B Drift velocity in a meridional and zonal pattern. It monitors seven ions, H^+, He^+, N^+, O^+, NO^+, N_2^+, and O_2^+. From those ions, SAMI3 evaluates the plasma/chemical composition of the ionosphere between 85 km to 20,000 km. Of those seven ions, H^+, He^+, and O^+ are used to solve ion/electron temperature equations. SAMI3 is able to output ion density, ion temperature, ion velocity, electron temperature, Total Electron Content (TEC), as well as determine the N_mF_2 and h_mF_2 (Garcia-Sage, 2019d; Huba et al., 2008).

CTIPe was developed by Timothy Fuller-Rowell et al. in conjunction with NOAA SEC (National Oceanic and Atmospheric Administration Systems Engineering Center). It is a physics-based model determining the ionosphere on a global scale. CTIPe utilizes the Weimer ionosphere electrodynamics model for calculating electric fields from solar wind density, solar wind velocity, and IMF (Interplanetary Magnetic Field) magnitude. It also employs a fixed/time-dependent Hemispheric Power Index. CTIPe is comprised of four components which include a global thermosphere model, a high-latitude ionosphere model, a mid and low-latitude ionosphere/plasmasphere model and an electrodynamical calculation. The "Coupled" part of the name is a result from the four components running concurrently in their utility towards solving continuity, momentum, and energy equations. As previously mentioned, CTIPe works on 3-dimensions with grid point resolution of 2° latitudinally, 18° longitudinally, and a vertical resolution depending on the pressure profile. CTIPe contains 15 distinct levels over a logarithmic variation of the column starting at 80 km. With this vast

resolution, CTIPe takes several parameters into account when solving the equations of motion. These parameters include Coriolis effects, horizontal pressure gradients, ion drag, and horizontal/vertical viscosity. With regards to the continuity equation, the major species for transport, chemistry and diffusion are O, N_2, and O_2 (Codrescu et al., 2008; Garcia-Sage, 2019a).

USU-GAIM was developed by R.W. Schunk, L.Scherliess, J.J. Sojka, D.C. Thompson, and L. Zhu in conjunction with the Center for Atmospheric & Space Sciences from Utah State University. It is a physics-based model that assimilates real-time measurements from a Kalman filter which makes it comparable to a Gauss-Markov Kalman Filter (GMKF) model. Working on a global scale, GAIM works with an Ionosphere Forecast Model (IFM) which covers altitudes of 90 km to 1400 km encompassing the E-region, F-region, and topside. Inputs for USU-GAIM include both F10.7 values, a daily Ap-Index as well as eight 3-hr Kp indicies. It also takes in data from up to 400 GPS ground stations worldwide. GAIM takes five ions (H^+, O^+, NO^+, O_2^+, and N_2^+) into account for its equations. The main output for GAIM is a 3-dimensional representation of the electron density profile, while additional outputs include N_mE, h_mE, N_mF_2, and h_mF_2 as well as slant/vertical TEC. By virtue of being a global model, the interplay between better spacial resolution and larger coverage results in a 15° longitude, 4.66° latitude grid box. With this limited resolution, smaller scale features including plasma bubbles and spread-F cannot be captured. Through CCMC, GAIM will assimilate GPS and TEC data between ±60° latitude (Garcia-Sage, 2019f; Schunk et al., 2004).

TIE-GCM was developed by R.G. Roble et al. in conjunction with the HAO NCAR (High Altitude Observatory, National Center for Atmospheric Research). It is a 3-dimensional physics-based model that is derived from first principles which couples the thermosphere and the ionosphere. The model begins calculations with a

few assumptions; hydrostatic equilibrium, constant gravity, steady-state ion/electron energy equations and isobaric incompressibility. With regards to first principles, TIE-GCM solves continuity, momentum, and energy equations for the following neutrals and ions: O^+, O_2^+, NO and NO^+. This is done with fourth-order, semi-implicit finite difference schemes for a grid of pressure surfaces. Its vertical extent ranges from 97 km to 500 km over 29 isobaric levels. The intervals for these levels varies by ½ scale height. With respect to latitude/longitude, this model uses a 5° x 5° grid. From an E×B drift, ion velocities can be determined. As inputs, TIE-GCM uses both F10.7 values and a 3-hr Kp index. It also employs a Heelis model, which requires cross polar cap potential, and a Weimer model, which requires the IMF, B_y, B_z, solar wind speed, and solar wind density. The outputs yield height of pressure surfaces, ion/electron/neutral temperatures, zonal/meridional/vertical wind speeds, chemical composition, electric potential and ion/electron densities (Garcia-Sage, 2019e; Richmond et al., 1992).

GITM was developed by A.J. Ridley et al. in conjunction with the Department of Atmosphere, Oceanic and Space Sciences out of the University of Michigan. Using a 3-dimensional spherical code, it is a physics-based model that emphasises a modern advection solver in order to use the continuity, momentum, and energy equations. The grid used to determine the thermosphere and ionosphere systems is stretched in both latitude and altitude with flexible grid points. Note that GITM uses an altitude grid as opposed to a pressure grid resulting in grid spacing 3 km in the lower thermosphere and 10 km in the upper thermosphere. The range in altitude goes from 90 km to 600 km. The flexible grid points as somewhat restricted through CCMC where the latitude/longitude resolution is 2.5° and 5° respectively. By virtue of covering all latitudes, the auroral region is covered by GITM. Whereas other models hold higher focus on ions, GITM explicitly solves neutral densities such as O, N, NO, O_2, and

N_2. GITM still takes ions into account as well including O^+, N^+, NO^+, N_2^+, and O_2^+. The model also differs by using a non-hydrostatic solution for realistic dynamics in the auroral zones. With each neutral species having a different vertical velocity, the frictional term couples the different velocities for more accurate relationships and shear results. GITM solves the ion momentum equation by assuming a steady-state and taking gravity, pressure, neutral winds, and electric fields in account. As inputs, GITM uses F10.7, IMF, solar wind velocity, and solar irradiance. The outputs are ion/electron/neutral temperatures, zonal/meridional/vertical neutral wind and plasma velocities, mass densities, and ion/electron/neutral number densities (Garcia-Sage, 2019b; Ridley et al., 2006).

IRI was developed by Dieter Bilitza at NASA GSFC. As part of an international project, this model was originally formed by a working group in the late 1960s sponsored by the Committee on Space Research (COSPAR) as well as the International Union of Radio Science (URSI). It was established as an emperical (climotological) model. For inputs, IRI uses daily at averaged F10.7 values (averaged over 81-days and 12-months), in addition to a running mean sunspot number. It also takes a daily and 3-hr Ap indicies as well as an ionosonde-based IG index (an alternative to sunspot number when f_oF_2 is being calculated). The altitude range for IRI extends from 50 km up to 2000 km. Through all of these altitudes, ion compositions (O^+, H^+, He^+, NO^+, and O_2^+) and other outputs, listed below, are tracked. The outputs for this model are electron density, ion/electron temperature, TEC, and ion compositions. Some of the major sources for climotological data feeding into this model include a worldwide ionosonde network, scatter radars (Arecibo, Jicamarca, Millstone Hill, St. Santin, and Malvern), topside sounders, and in situ instruments via rockets and satellites. By virtue of it being a climotological model, it is an ideal source for providing monthly non-auroral ionosphere values with a high degree of confidence during magnetically

quiet conditions. We are using IRI-2016, but the model is updated annually following workshops and general assembly meetings. Additional extensions are expected with new features including ion drift, auroral & polar ionosphere calculations, and attempting to consider effects of magnetic storming (Bilitza et al., 2007; Garcia-Sage, 2019c).

2.2 Ray Tracer

Ray tracers have been used as a simplification of focusing or visualizing the path waves would take provided they behaved as a beam instead of a wave through the environment they are traveling though. The ray tracer used for this project, developed by AFRL, uses a modified Jones-Stephenson code (Dao et al., 2016). The ray tracer does assume Earth is perfectly spherical as opposed to the known truth that the Earth is indeed an oblate spheroid. This is a safe assumption to make as the distance between points in the WSMR experiment is only a couple degrees latitude/longitude.

As with most ray tracers, the basis for its functionality lies with the proper utilization of the Jones-Stephenson algorithm established in 1975 (Jones and Stephenson, 1975). The utility of the algorithm is strengthened when taking Earth's magnetic field into account and altering the ray tracing algorithm to consider Appleton-Hartree equations. Ultimately this allows an individual to consider the X-mode as well as the O-mode.

The ray tracer used in this process was written using Julia, a relatively new computer language. The main motivation for the AFRL designed ray tracer was to allow for a more user-friendly and customizable final product removing the necessity of FORTRAN knowledge which was the original language developed by Jones and Stephenson (1975). A major focus ensuring accuracy and compatibility with the

original source code was cross-checking all final results from Julia were comparable to Jones-Stephenson in regards to numerical differences and a magnetic field model. One key difference is the utilization of an up-to-date IGRF (International Geomagnetic Reference Field) magnetic field model.

The equations used by Jones and Stephenson (1975) for ray tracing work off of Hamiltonian equations which exist in spherical polar coordinate system. A group path variable, $P' = ct$, simplifies the equations as P' derivatives are independent to the choice of Hamiltonian. This variable was selected granting FORTRAN flexibility to switch Hamiltonians fluidly at any point in the path which results in incrementally smaller path length steps as you approach the point of reflection. In addition to P', Table 1 defines the selection of variables used to work through the ray tracing equations and Equations 1 - 6 show the equations numerically solved in a ray tracing program:

Table 1. Definitions of variables used in Jones and Stephenson (1975) ray-tracer.

Variable	Description
P'	Group path ($P' = ct$)
H	Hamiltonian
r, θ, ϕ	Spherical polar coordinates
c	Speed of light
k_r, k_θ, k_ϕ	Components of the propagation vector
ω	Angular frequency ($2\pi f$)
f	Frequency of Electromagnetic wave
t	Time

$$\frac{dr}{dP'} = -\frac{1}{c}\frac{\partial H/\partial k_r}{\partial H/\partial \omega} \tag{1}$$

$$\frac{d\theta}{dP'} = -\frac{1}{rc}\frac{\partial H/\partial k_\theta}{\partial H/\partial \omega} \tag{2}$$

$$\frac{d\phi}{dP'} = -\frac{1}{rc\,sin\theta}\frac{\partial H/\partial k_\phi}{\partial H/\partial \omega} \tag{3}$$

$$\frac{dk_r}{dP'} = -\frac{1}{c}\frac{\partial H/\partial r}{\partial H/\partial \omega} + k_\theta\frac{d\theta}{dP'} + k_\phi\, sin\theta\frac{d\phi}{dP'} \tag{4}$$

$$\frac{dk_\theta}{dP'} = \frac{1}{r}\left(\frac{1}{c}\frac{\partial H/\partial \theta}{\partial H/\partial \omega} - k_\theta\frac{dr}{dP'} + k_\phi r\, cos\theta\frac{d\phi}{dP'}\right) \tag{5}$$

$$\frac{dk_\phi}{dP'} = \frac{1}{r\, sin\theta}\left(\frac{1}{c}\frac{\partial H/\partial \phi}{\partial H/\partial \omega} - k_\phi\, sin\theta\frac{dr}{dP'} - k_\phi r\, cos\theta\frac{d\theta}{dP'}\right). \tag{6}$$

There was also focus put on ensuring the propagation vector would produce a normalized wave normal direction such that:

$$k_r^2 + k_\theta^2 + k_\phi^2 = \frac{\omega^2}{c^2} \tag{7}$$

A key component to the success of a ray tracer is the selection of the Hamiltonian which is allowed to alter during the path of the ray. In Jones and Stephenson (1975), the Hamiltonian can be altered under several different subroutines with respect to different conditions such as the presence of collisions or a magnetic field. These different subroutines are encoded AHWFWC, AHWFNC, AHNFWC, & AHNFNC to indicate it is derived an Appleton-Hartree equation (AH) with\neglecting a magnetic field (WF or NF) and with\neglecting collisions (WC or NC). An example of one Hamiltonian utilizing Appleton-Hartree equations looks like:

$$H = \frac{1}{2}\left(\frac{c^2}{\omega^2}(k_r^2 + k_\theta^2 + k_\phi^2) - real(n^2)\right) \tag{8}$$

where n is the refractive index.

A sample output for the ray tracing results is seen in Figure 1. The top-left image shows longitudinal variation during the ray's path as well as the peak altitude wheres the top-right shows the latitudinal component throughout ray path. The bottom-left image is a compilation of the top images as if from a "bird's eye view". The minor bend near 33° latitude is from the minor alteration in the path from the top-left image. The colors in the images correspond to the plasma frequency observed at a given altitude. This allows users to see plasma bubbles or other altitudinal variation that may interact with the ray. The bottom-right image is the EDP the ray runs through.

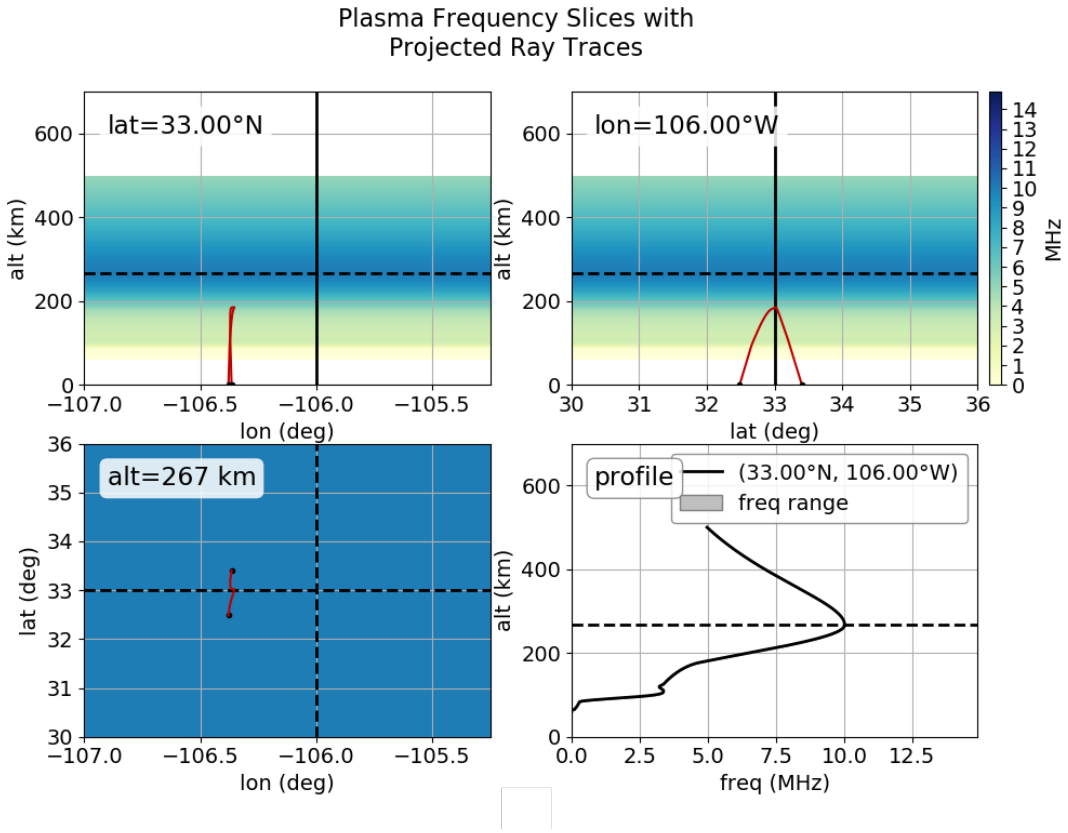

Figure 1. An example of the AFRL ray tracer output for a given EDP.

2.3 Similar Validations

Earlier, several examples were discussed with regards to previous validation efforts and alternate approaches to legitimizing a variety of ionospheric models. One of those examples included work done by Ja-Soon Shim of CCMC with her work on vertical drift among other items. This validation effort (Shim et al., 2011) was in association with the CEDAR (Coupling, Energetics, and Dynamics of Atmospheric Regions) ETI (Electrodynamics Thermosphere Ionosphere) Challenge in order to quantitatively assess prediction capabilities of particular ionospheric models. This particular validation effort was a driving mechanism to determine statistical parameters we wish to emulate and allow for any future implementation of this technique.

To validate the results in this ETI Challenge, root-mean square (RMS) was used to compare observed and model-derived data.

$$RMS_{mod} = \sqrt{\frac{\Sigma(x_{obs} - x_{mod})^2}{N}}. \tag{9}$$

The basic RMS_{mod} equation calculates the difference between your observed (x_{obs}) and model-derived (x_{mod}) values with N representing the number of values you're comparing. If you were to get an RMS_{mod} of 0, this indicates a perfectly matching performance by the model (Shim et al., 2011).

Prediction efficiency (PE) is another statistic used in Shim et al. (2011) for the purpose of evaluating a model's performance:

$$PE = 1 - \frac{RMS_{mod}}{RMS_{ref}} = 1 - \sqrt{\frac{\Sigma(x_{obs} - x_{mod})^2/N}{\Sigma(x_{obs} - \langle x_{obs} \rangle)^2/N}}. \tag{10}$$

The RMS_{ref} is a reference value of observed data which is ultimately the variance

or deviation from a mean value $\langle x_{obs} \rangle$. When PE = 1, there is a perfect prediction. When PE = 0, the error of the model is comparable with the variation of observed values. Any point where PE < 0 lets the user know the model error was large and unreliable because variation in the model was larger than observed variation (Shim et al., 2011).

One final metric from the CEDER ETI Challenge (Shim et al., 2011) is the ratio of maximum change which is a measurement determining how closely the peaks and valleys of models and observations matched. Whereas RMS and PE can be used to long-term calculations, the ratio can better show shorter temporal variations over any time interval:

$$Ratio\ Range_{(max-min)} = \frac{(x_{mod})_{max} - (x_{mod})_{min}}{(x_{obs})_{max} - (x_{obs})_{min}}. \tag{11}$$

With RMS, PE, and Ratio Range as parameters, a complete quantitative picture can be formed with how well a model may have done compared to a known observed value. These metrics were all applied to a larger validation effort meaning that it covered multiple latitudes over a longer period of time as opposed to the singular location explored through this proof-of-concept.

Another document discussed was a high resolution model known as GPSII which was developed by Nickisch et al. (2015). GPSII works similarly to GAIM were it relies on direct data assimilation, but a major difference is that GPSII is not treated as a global model which allows it to fine tune its area of interest. GPSII is able to accurately map out the ionosphere even during storming conditions because of its temporal and spatial resolution. With regards to assimilation, the data sources GPSII ingests are ionogram data and GPS data (Nickisch et al., 2015). With use of local ionograms and beacon data, GPSII excels as a fine resolution, local ionospheric model.

GPSII was the model chosen by another IARPA-funded project looking specifically

at HF Geolocation which intentionally utilized data containing medium-scale TIDs. To enhance accuracy, GPSII was adjusted to include delay-Doppler data as well as implementing AoA information from known locations identified as known reference points (KRPs) (Nickisch et al., 2015). These KRPs are the exact same locations from Figure 2 to be discussed in further detail in Section 2.4.

Nickisch et al. (2015) lays out the results of the error in terms of degrees East of Zenith and degrees North of Zenith after displaying the measured results from the IARPA experiment and the results of a Jones-Stephenson ray tracer which designates GPSII's results. They take the directional errors and combine them to a singular scatter plot. Finally, they developed a Cumulative Distribution Function (CDF) showing a solid angle error which is a recasting of the directional errors. The CDF is a graphical display of the required error to obtain a percentage of the total data points. For example, in Nickisch et al. (2015), their Figure 3 reveals that in order to obtain 80% of the results, you would have to extend out to 1 mSR of error, but to encompass all data points (100%), you have to extend out to 9 mSR. Through these four plots, we are able to expand upon our simple error measurements and reshape them to an application-focused result.

The work done by Nickisch et al. (2015) is deemed as a novel effort regarding GPSII's accuracy and ability to model AoA data, time delay, and Doppler fluctuations during medium-scale TIDs. The model's success is attributed by the author to the "power of Doppler shift as a measure of TID structure and corresponding ionospheric tilts" (Nickisch et al., 2015). The ability for this model to remain 90% accurate during TIDs gives promise to other future modeling efforts. Not only was GPSII able to model a turbulent ionosphere with tremendous accuracy, but the simplistic process which produced significant data is the basis of our project.

2.4 WSMR Experiment

With the models covered in Section 2.1 and the ray tracer detailed in Section 2.2, someone is capable of comparing the outputs of the models to each other, but those results are meaningless until you relate them to some ground-truth data as Nickisch et al. (2015) did for GPSII. The ground data was provided in January 2014, when IARPA funded an HF geolocation experiment which occurred in White Sands Missile Range in New Mexico. Figure 2 displays the layout of this experiment. Over the course of a 9 day period, a variety of frequencies were transmitted from multiple transmitters to a single receive site.

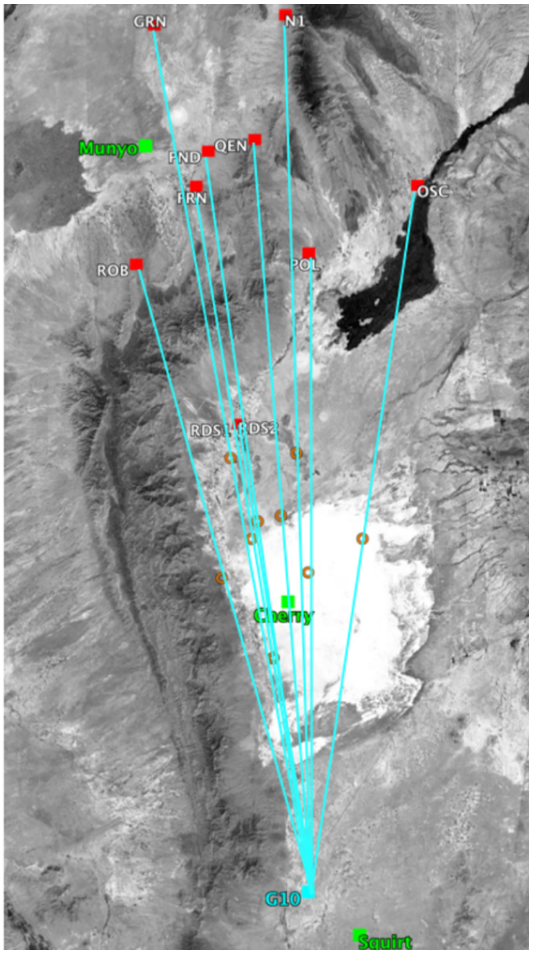

Figure 2. Layout of WSMR experiment with 8 transmitters to the North and one receiver to the South (G10).

For the purpose of this thesis, we are focusing on one particular day of the experiment when a designated frequency (5.33 MHz) was transmitted from 6 transmitters (ROB, GRN, FRN, PND, POL, and QEN) and a single receiver (G10) captured the signal. From this point, AoA data (azimuthal angle and elevation angle) and group delay information was gathered. Figure 3 is a visualization of AoA where the azimuthal angle would indicate the bearing or direction the signal came from and the elevation angle would determine the steepness of the signal from zenith. Group delay (group length / speed of light) is essentially an integrated group path which measures the time between transmission and reception. Azimuthal and elevation angles can be simplified into one solid Angle of Arrival angle and then recast as a solid angle which will be explained in better detail later.

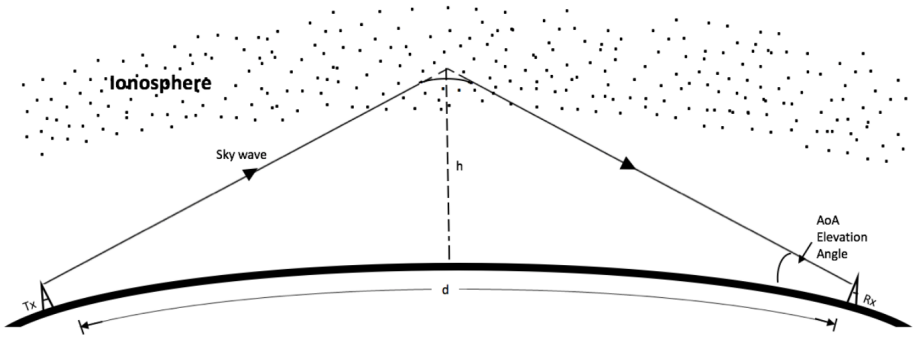

Figure 3. A visual representation showing AoA. (McNamara, 1991).

The data collected during this experiment was the "ground truth" which can then be compared to any model derived data. One means of comparison is analyzing Electron Density Profiles (EDPs). An ionogram, like the one in Figure 4, reveals the EDP the signal was sent through during the Jan 22 experiment.

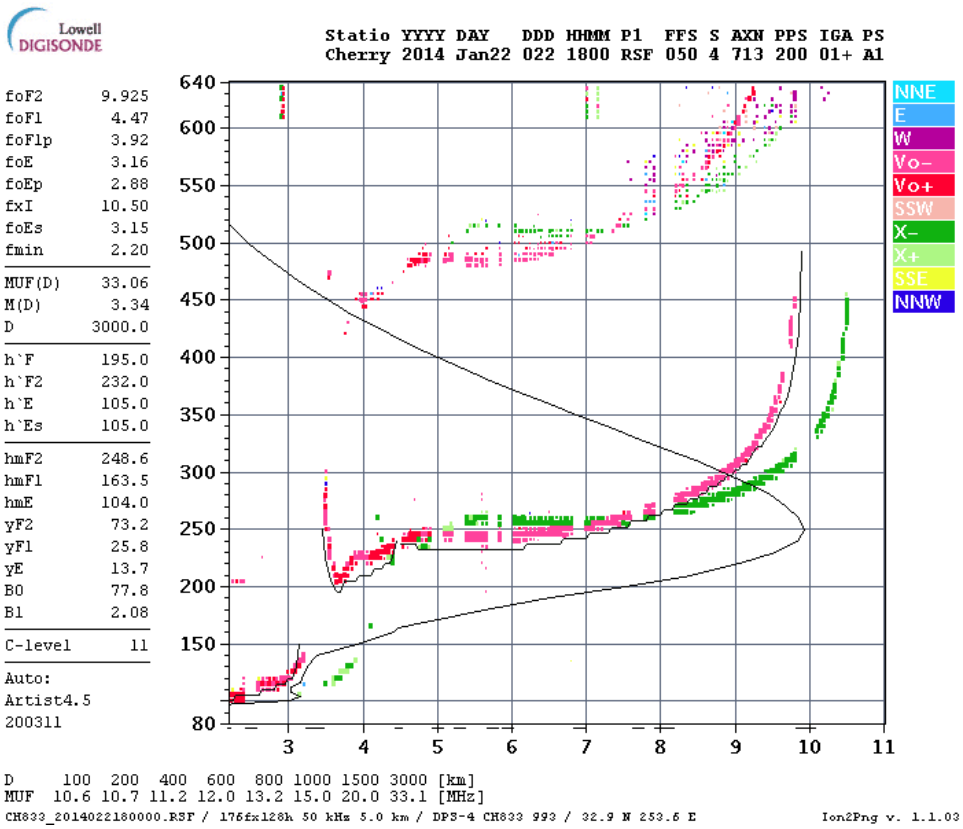

Figure 4. Visual representation of the state of the ionosphere over WSMR at 1800Z on 22 Jan, 2014 as presented by an ionogram.

If the ionosphere were a simple planer surface, the mental picture of a signal bouncing off a mirror with a basic reflection/refraction reaction would be adequate comparable to Figure 3. This basic set-up is known as the "Mirror Model" (McNamara, 1991) and though it is a reasonable facsimile to what is going on, it is not always the most accurate visualization of what is occurring. In the event of Traveling Ionospheric Disturbances (TIDs), one could just as easily describe the ionosphere as dynamic and fluid with grooves, dimples, and divots like a golf ball resulting in large deviations in time aloft and the path a beam would take.

III. Methodology and Results

This chapter will be a step-by-step guide through the metrics, results, and rationale for a validation effort using AoA data. We will also be exploring how this technique can apply to the X-mode, in addition to the O-mode, which essentially doubles your available data. It is important to know that the results you are about to see are from one day, with one frequency, over one area. We are establishing a methodology one may follow and implement throughout their validation efforts

3.1 Data collection

As with any validation, measurements must be collected in order to have values to compare against. A "ground-truth" set of information gathered through reliable means allows for increased confidence that you have promising data for comparison. As discussed in Chapter 2, we used the information gathered from the WSMR experiment to represent the "true state" of the ionosphere over the region during that time. For the purpose of our representative validation, we ensured the selected data occurred during a time of no significant geomagnetic storming, no TIDs, and a continuous source of available data. For example, Figure 5 shows an 8 hour time frame, but there is a feature moving in around 22Z where the variability in azimuth angle is wide sweeping ranging from -38° to 12°. Figure 5 also shows a small period of time near 19Z with a slight data gap, but overall the data is continuous. These guidelines were established early on so we may later give the models the best chance for success. In our example, we had a 5.33 MHz signal transmitted through 6 different transmitter/receiver links due to the fact that these 6 were the ones with available data (no extended periods of time where data gap occurred). The signal was sent to the receiver where group path data and AoA (Angle of Arrival) data, which is comprised of

azimuthal angle and elevation angle, were obtained. Using an integrated group path, one can later determine a group delay in order to measure a time relationship to the distance the signal was aloft. The azimuth/elevation information can be a method for displaying spatial information revealing the direction of the incoming signal (with 0° azimuth representing North) as well as the steepness of the signal (with 90° elevation representing zenith). The group delay is a tool for temporal information as it shows the passage of time between transmission and reception. Once the data is received, a simple plot-over-time, like those seen in Figures 5, 6, and 7, can show the variability of the ionosphere even during quiet ionospheric conditions.

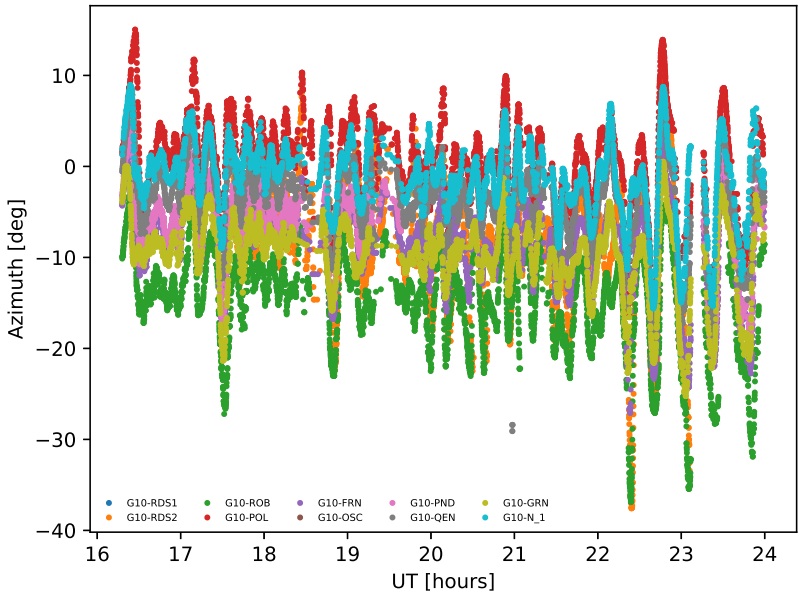

Figure 5. The received azimuth data for 5.33 MHz frequency on 22 Jan. The different colors represent the different transmitter-receiver links.

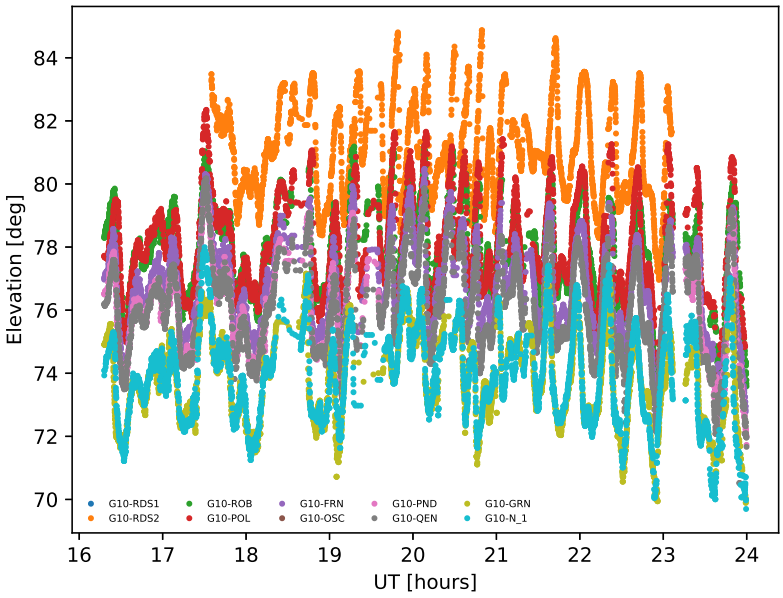

Figure 6. The 22 Jan elevation data for 5.33 MHz. Contrary to Figure 5, 22Z lacks the apparent feature seen on Figure 5 which may be a tilt in the ionosphere.

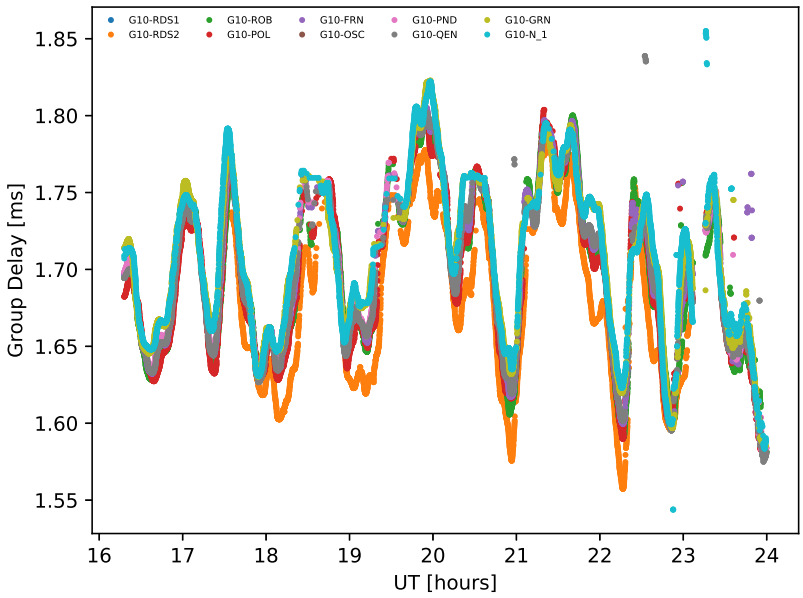

Figure 7. The 22 Jan group delay data for 5.33 MHz. The variation between each link is minimal compared to variations of AoA parameters.

Each of these figures displays a highly variable ionosphere despite occurring during "quiet" conditions. The different colors within each plot also shows the different links over time. This is the ground truth we will be comparing the models against. We will not be comparing results to the entire 8 hours, but rather just the 18Z - 20Z time period. In order to perform the comparison, we need to establish the models' interpretation of the ionosphere and that is where CCMC comes into play. Through CCMC, you are able to perform runs of ionospheric models and look for specific criterion including electron density, ion composition, etc. We will be using requested runs of the six models described in Section 2.1: TIE-GCM, CTIPe, SAMI3, GITM, USU-GAIM, and IRI. Following the data gathered from model output, you can compare against the measured EDP.

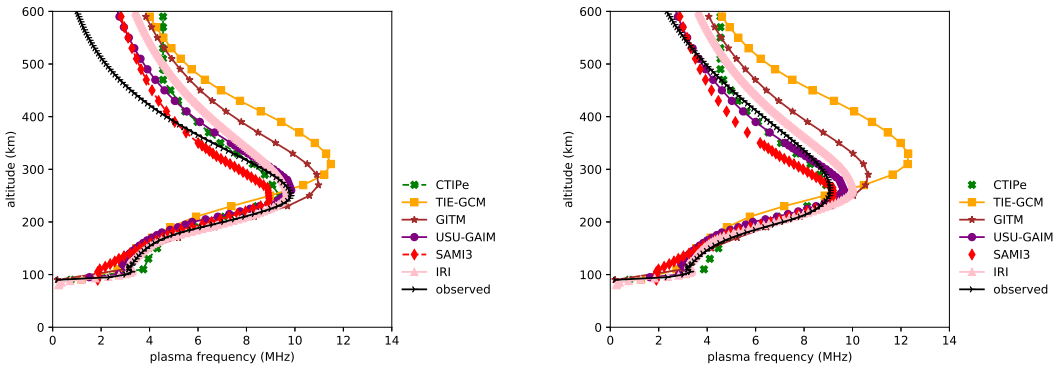

Figure 8. Meaured and modeled electon density profiles for 18Z (left) and 19Z (right) on 22 Jan.

A quick glance at Figure 8 shows that even given similar date, time, and latitude/-longitude information, models can vary greatly with regards to plasma frequency as a function of altitude. The most obvious difference is between the observed F2 peak (f_oF_2 at 9.9 MHz, h_mF_2 at 248.8 km) and the peak for TIE-GCM (f_oF_2 at 11.5 MHz, h_mF_2 at 310 km)in addition to the drastic difference of the E-region for CTIPe and that of all other models. These plots inherently have different smoothness to them as a result of the altitude resolution.

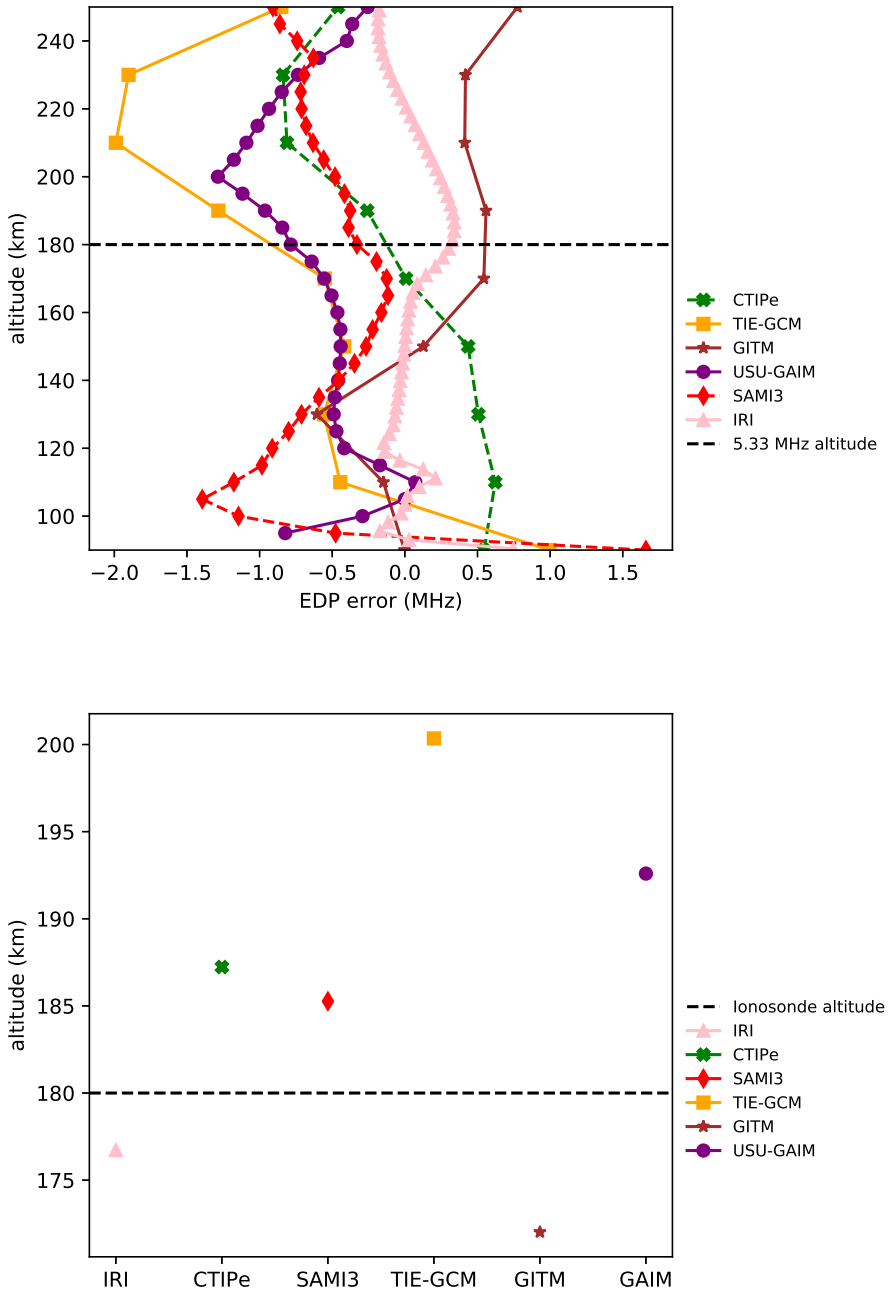

Figure 9. (Top) The altitude of the first observed value to the observed F2 peak, the difference between the model and observed values helps display where future error may come from. The dashed line denotes the altitude where 5.33 MHz meets the observed EDP. (Bottom) Each model's altitude value when intersecting 5.33 MHz.

Figure 9 depicts two different visualizations of potential sources for error. The top image is a zoomed in view of the bottomside ionosphere focusing on the difference in plasma frequency as you increase in altitude. Here, the difference between the model and observed frequencies from the 18Z plot of Figure 8, is accenting whether a model overestimated/estimated the plasma frequency at given altitudes. For example, paying close attention to CTIPe, the model overestimates the plasma frequency between the altitudes of 90 km to 170 km. The dashed line at altitude 180 km is the altitude where the 5.33 MHz signal intersects the observed EDP. Looking at this altitude and those below it represent that area of the ionosphere the signal would interact with before being reflected towards the surface. It is important to consider the relative error. Though the error in frequency is "only" 1 MHz, this relates to a nearly 50% error in some locations whereas a 1 MHz error near the F2 peak only relates to a 10% error. The bottom plot describes the spread between each model and where they individually intercept 5.33 MHz. Once again, focusing on CTIPe, the altitude where CTIPe intercepts 5.33 MHz is around 187 km which is higher than the observed value which is around 180 km. It stands to reason that a more elevated point of refraction would subsequently increase the model's group delay as well as decrease the elevation angle as the beam would be forced to travel longer and be picked up from a shallower angle. For both plots in Figure 9, cubic splines were used to interpolate between established values whereas the construction of Figure 8 employed a basic linear plot. This in turn could cause some deviation between each point if performing a simple eye-ball comparison of where the model intersects 5.33 MHz.

3.2 Metric Formulation

From here, we used a homing technique to find the correct path for each ionospheric profile and got results comparable to the Figures 5, 6, and 7.

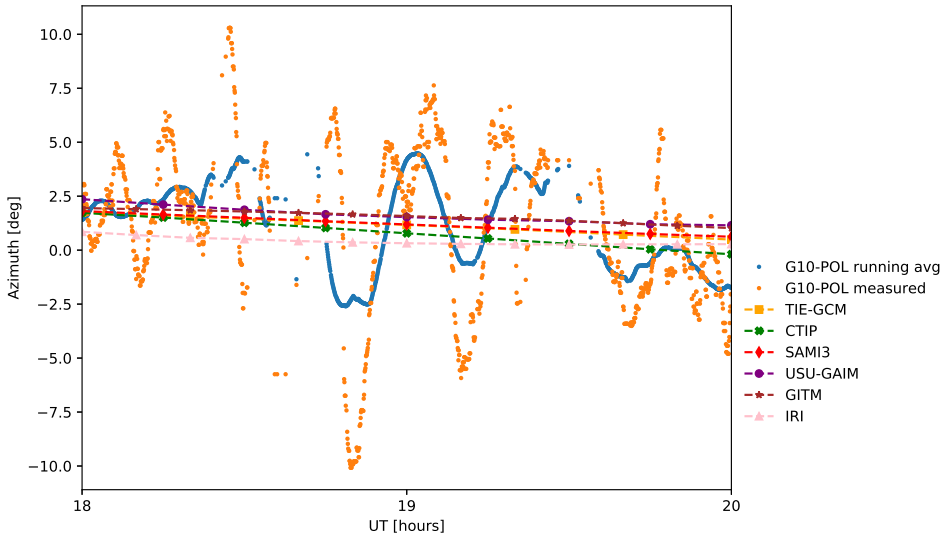

Figure 10. The modeled value of azimuth compared to measured values over time for POL-G10 link.

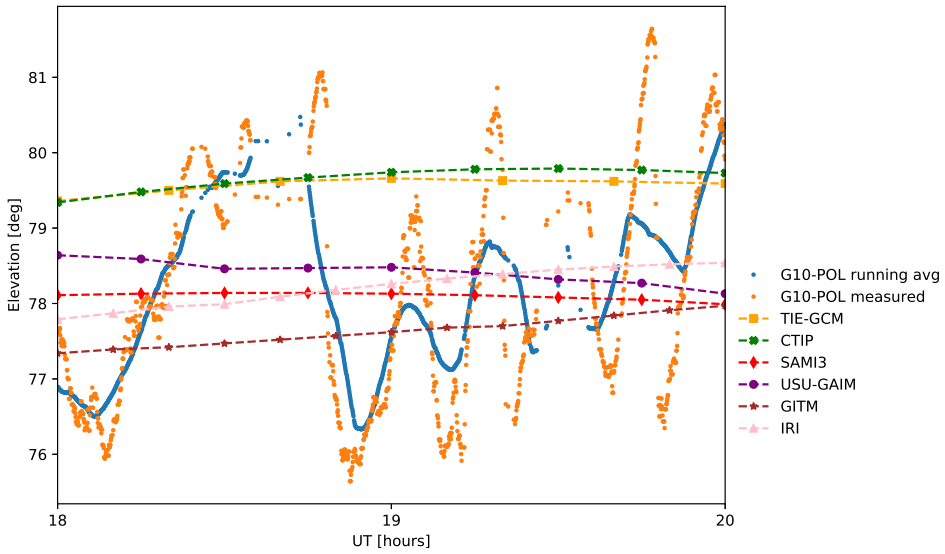

Figure 11. The modeled value of elevation angle compared to measured values over time for POL-G10 link.

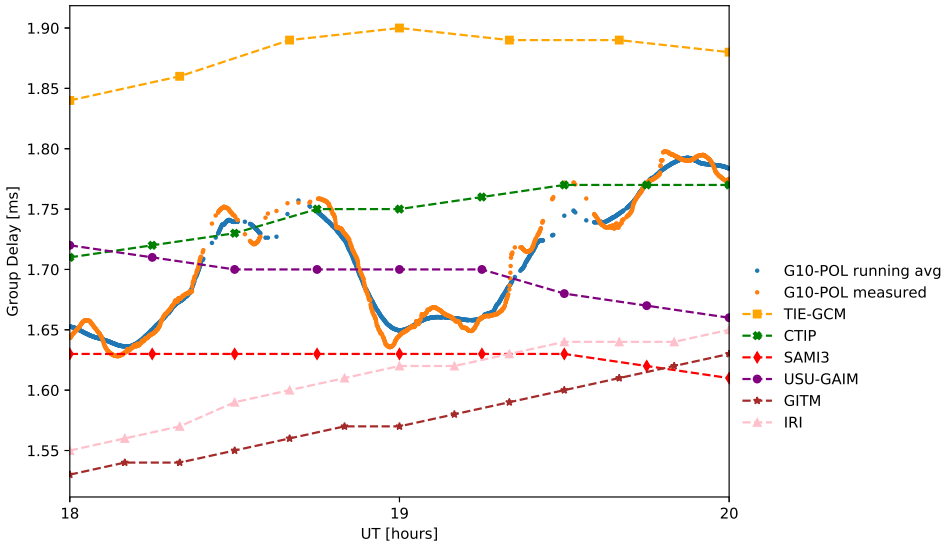

Figure 12. The model value of group delay compared to measured values over time for POL-G10 link.

For Figures 10, 11, and 12, the orange dotted line is the true measured results from the experiment, but note the instances of sparse data (i.e. around 1845Z) and the large variability in the measurements. With these data points capable of causing future issues, as they restrict utility towards calculating metrics, we have established a 10-minute running average in order to smooth out the data, eliminate outliers, and minimize gap potential. The uneven spacing for data points between the models is a result of the time resolution for each model where TIE-GCM has a time resolution of 20 minutes, CTIPe/SAMI3/USU-GAIM have a time resolution of 15 minutes, and GITM/IRI has a time resolution of 10 minutes. With these plots, a few details come clear. First, notice that despite high time resolution of the observation resulting in massive swings in values, the models' values remain in the range of observed values. Secondly, trends become apparent over time. For example, examine the general path of the observations in Figure 12 where the trend is ascending over time whereas the path of USU-GAIM appears to descend. Additionally on Figure 12, you can see the relationship between a different altitudes on the EDP for 5.33 MHz and different

delays (higher altitude, longer delay) which helps emphasize the point that a ray would have to travel farther/longer between transmission and reception.

To further emphasize how well the models compare to the observations, we can compile statistical measurements including their maximum/minimum/average values. In Figures 13, 14, and 15, the gray background displays the bounds of the observed maximum and minimum while the dashed line represents the average value. Each model has a similar set-up with tack marks designating their maximum and minimum values and a marker displaying their average value.

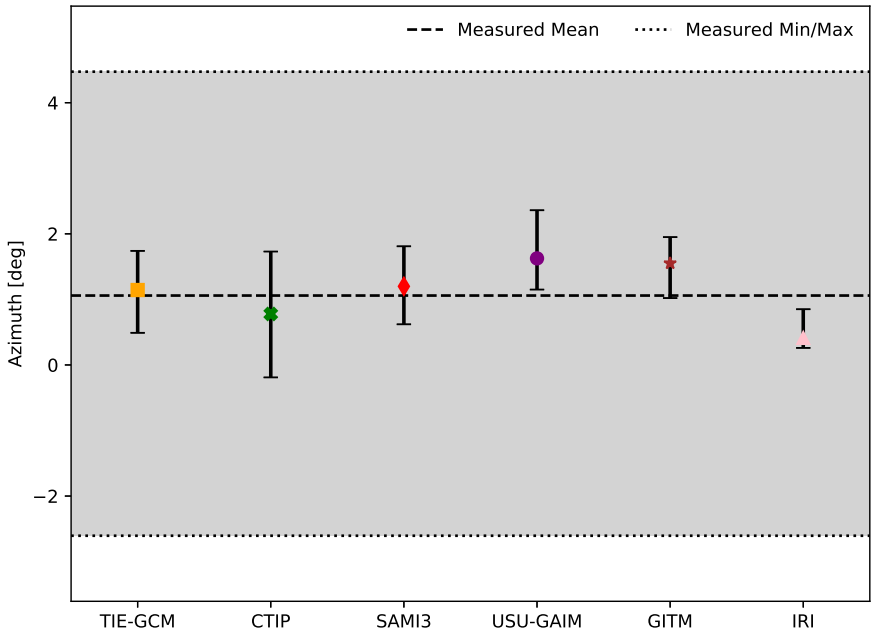

Figure 13. Maximum/minimum/average azimuth values of observations and each model for POL-G10 link. Each model was near the average value of the observations and their maximum/minimum values were well within the extremes of the observations.

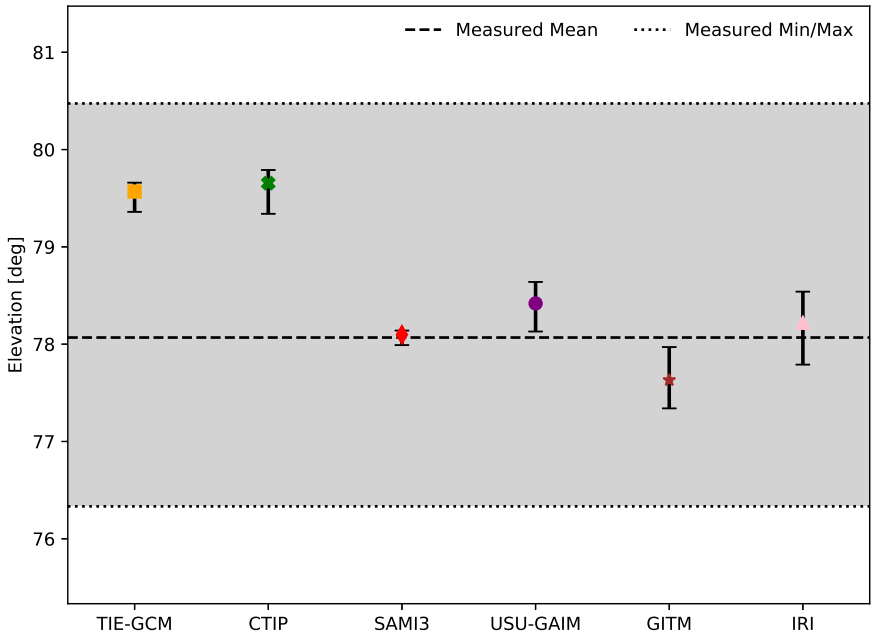

Figure 14. Maximum/minimum/average elevation values of observations and each model for POL-G10 link. The models had lower variance between their maximums and minimums. There is a larger spread between the models' average elevation angle compared to azimuth angle.

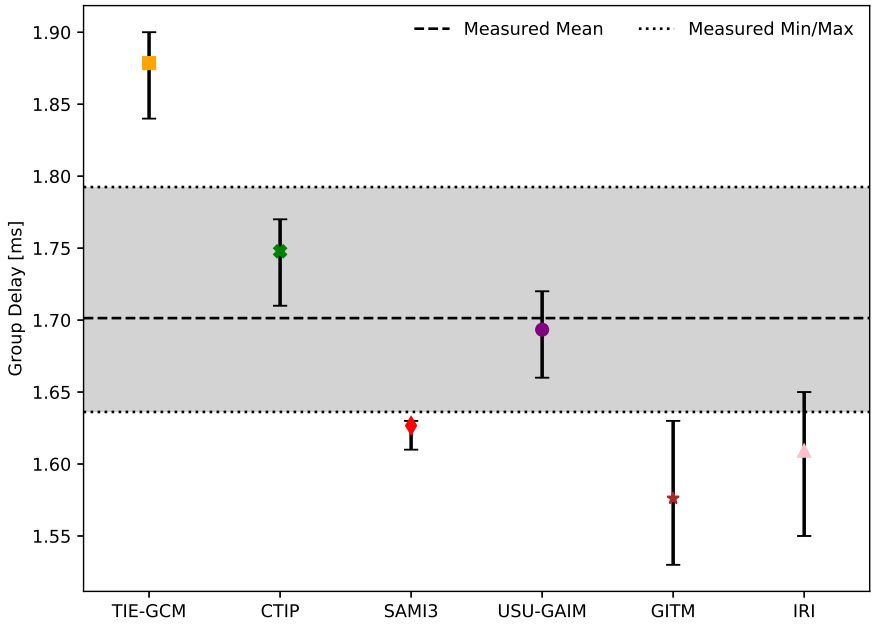

Figure 15. Maximum/minimum/average group delay values of observations and models for POL-G10 link.

After you have a comparison between models and observations, you can start performing metrics and additional statistics. A quick metric is error which is calculated as:

$$Error = x_{mod} - x_{obs} \tag{12}$$

where x_{mod} and x_{obs} are the model and observed values of whatever parameter you are comparing at that time. Figures 16, 17, and 18 show this error over time where the error is the model value at a particular time subtracted from the running average value at that time. The different time resolutions are more evident in these plots as TIE-GCM (one observation every 20 minutes) isn't as smooth in transitions as IRI (one observation every 10 minutes).

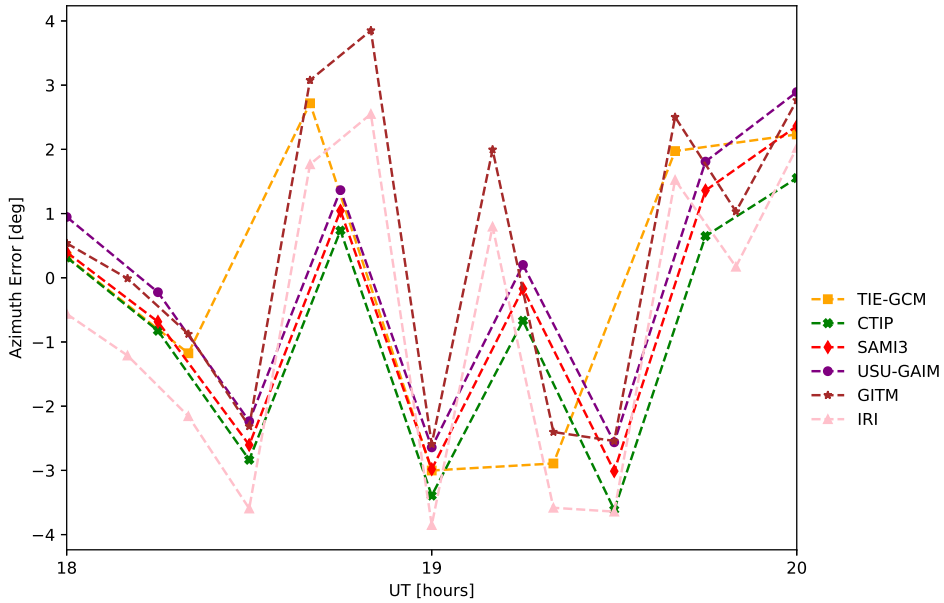

Figure 16. Deviation of models from observed azimuth value for POL-G10 link. The observed value is the running average which is the average value within a 10 minute window.

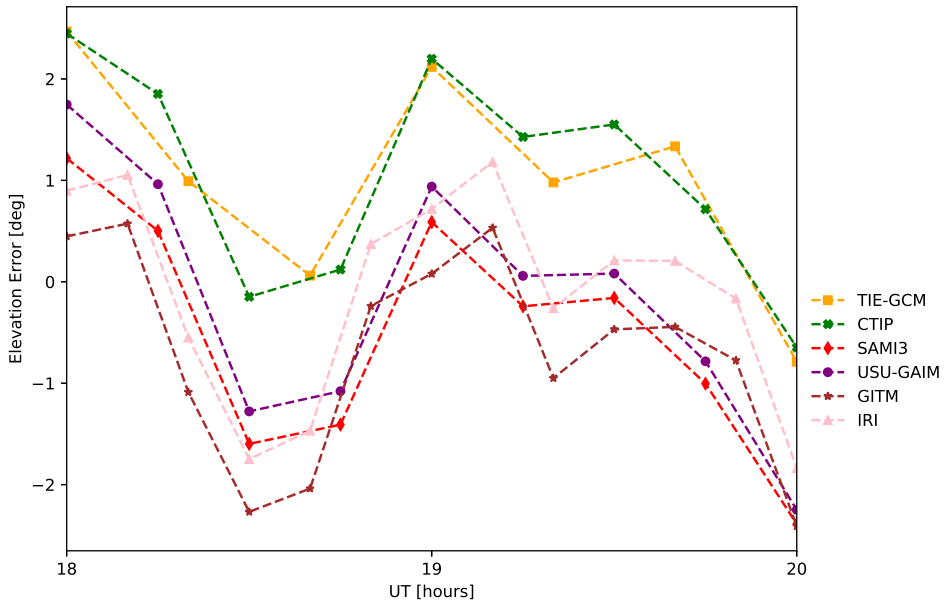

Figure 17. Deviation of models from observed elevation value for POL-G10 link.

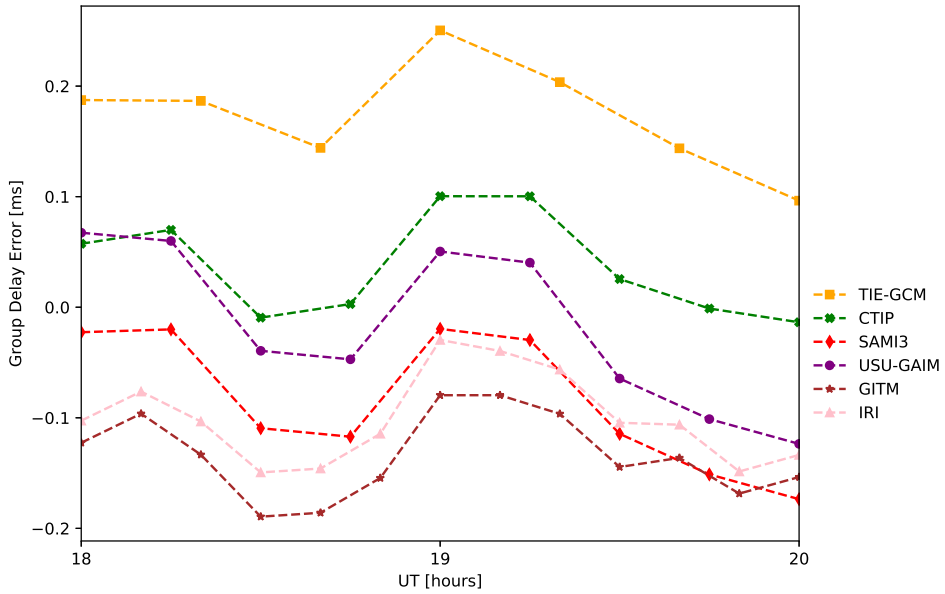

Figure 18. Deviation of models from observed group delay value for POL-G10 link.

There are a couple items to point out. In Figure 17, it could be seen that the models that started off too high in Figure 11 ended up with the least error by the end of run whereas the models who nearly matched the observations early on had a

larger error near the end. This is a clear representation of the mostly stagnant model values compared to the high variation over time of the observations. In Figure 18, one big takeaway is though USU-GAIM appears to hover around the 0.0 error mark and almost perfectly match the average value on Figure 15, it is important to remember that the trend seems to be a descending trend over time on Figure 12.

From this point, Mean Absolute Error (MAE), as opposed to Root Mean Square Error (RMSE), is the metric used to quantify the error. This was an intentional deviation from other works including Shim et al. (2011) because of the willful attempt to minimize the influence of the extremes. The error is calculated as deviation from a rolling average which is the smoothed out curve formed by taking a calculation window (in this case, we used a window of $\pm$ 5 minutes) and taking an average of the values. If one is willing to accept extremes in their calculations, the practice of taking RMSE (Equation 9) allows for extra weight given to any outliers from squaring the error. The equation for MAE, when taken in conjunction with the "observed" value being the rolling average, eliminates that bias towards extremes. It simply takes the error for face value by:

$$MAE = \frac{\Sigma(|x_{mod} - x_{obs}|)}{N} \tag{13}$$

where x_{obs} and x_{mod} are the model and observed values of whichever parameter is being analyzed and N is the number of observations. Even with a rolling average for RMSE, an intrinsic amplification towards the error is still evident. The error is exacerbated for the models' lower time resolutions. The low time resolution problem also is a concern for metrics. Focusing on one single link will not be "statistically significant" only having 7, 9, or 13 observations depending on which model is in use. The MAE data received per link is still capable of revealing some insight, though the results should be approached with some caution. The problem of being "statistically significant" will be corrected near the end of the process when we combine all links.

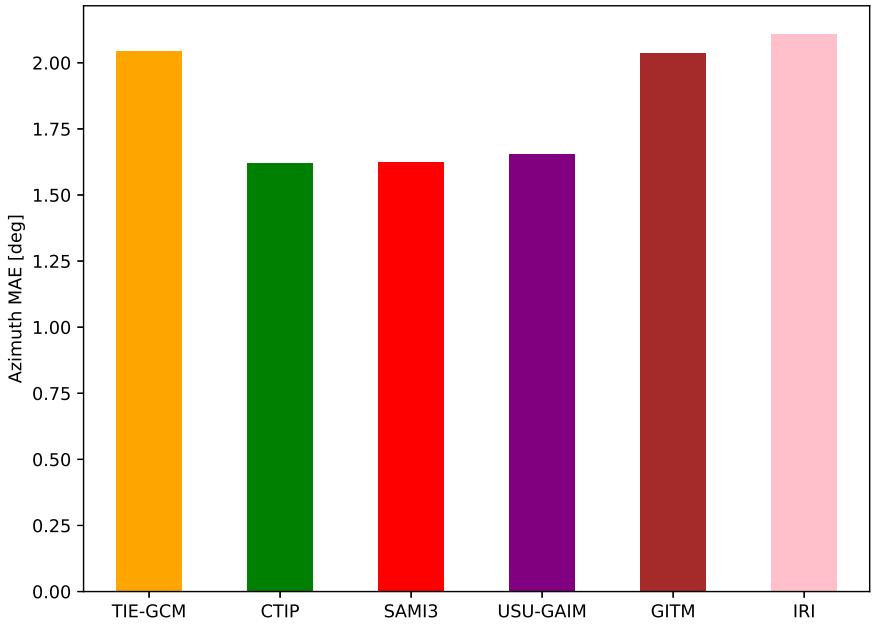

Figure 19. Azimuth MAE for the POL-G10 link shows time resolution does not guarantee less error. IRI and GITM have larger MAE than CTIPe despite having more observations per hour.

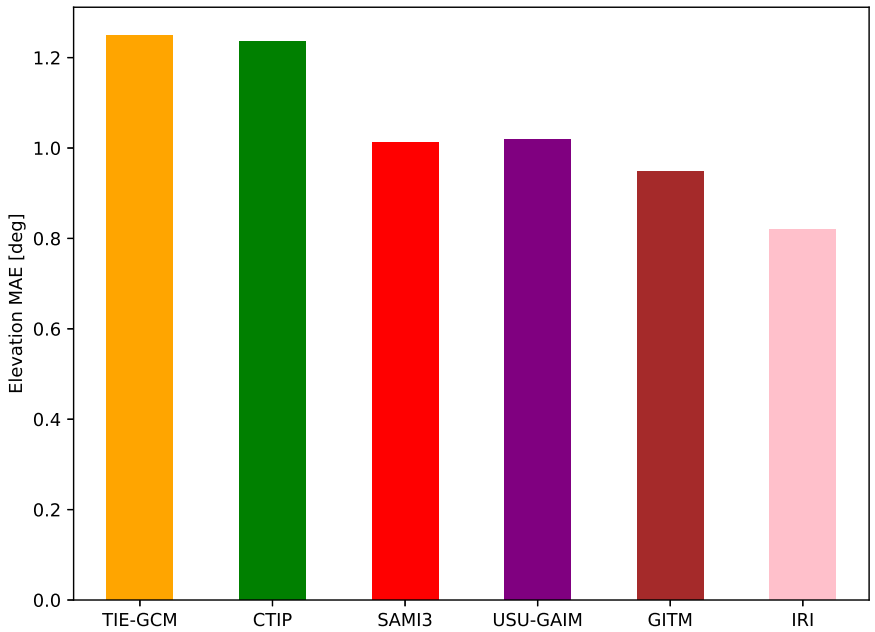

Figure 20. Elevation MAE for the POL-G10 link reveals IRI and GITM performed much better when compared to their respective azimuth error. This leads to the idea that having two spatial parameters may not be a conclusive measurement.

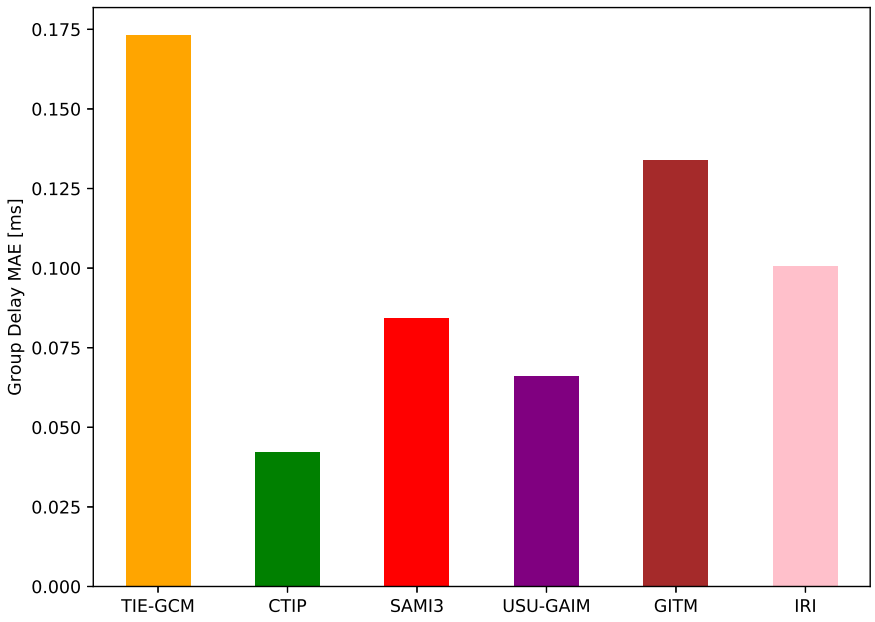

Figure 21. Group delay MAE for the POL-G10 may show a "small" error with 0.100 ms, but that relates to a mistaken distance of about 30 km which is of importance for HF geolocation applications.

Some entities may feel that determining the error and MAE of the models' performances may be enough for a validation effort because it quantitatively displays the deviation of the model from the observed value. If a model did well, the error and MAE would be consistently low, but just like football teams can win on "any given Sunday", we want to expand the validation to ensure that we don't prematurely deem a model as a valid tool. For example, we see with regards to azimuth angle (Figure 19), CTIPe and SAMI3 performed well with USU-GAIM not too far off, whereas IRI was the least successful model for this particular link, on this particular day. With regards to elevation angle (Figure 20), almost the opposite is true with IRI doing well whereas CTIP was among the poorer performers. When looking at the more temporal parameter of group delay (Figure 21), CTIP was once again good, but TIE-GCM misrepresented the state of the ionosphere. With three different results for three different parameters, how could a "best/worst" possibly be selected. This

implies we need additional metrics and possibly a different condensed measurement to paint a clearer picture.

Ultimately, we are trying to see how well a model matched the data received from a known transmitter. Numerically determining error can be of some value, but there is another way to display the error. We can graphically display each models' rays showing the AoA errors, azimuth and elevation, on a radar plot. This radar plot will take the previously determined errors and compare them to the known reference point of the transmitter.

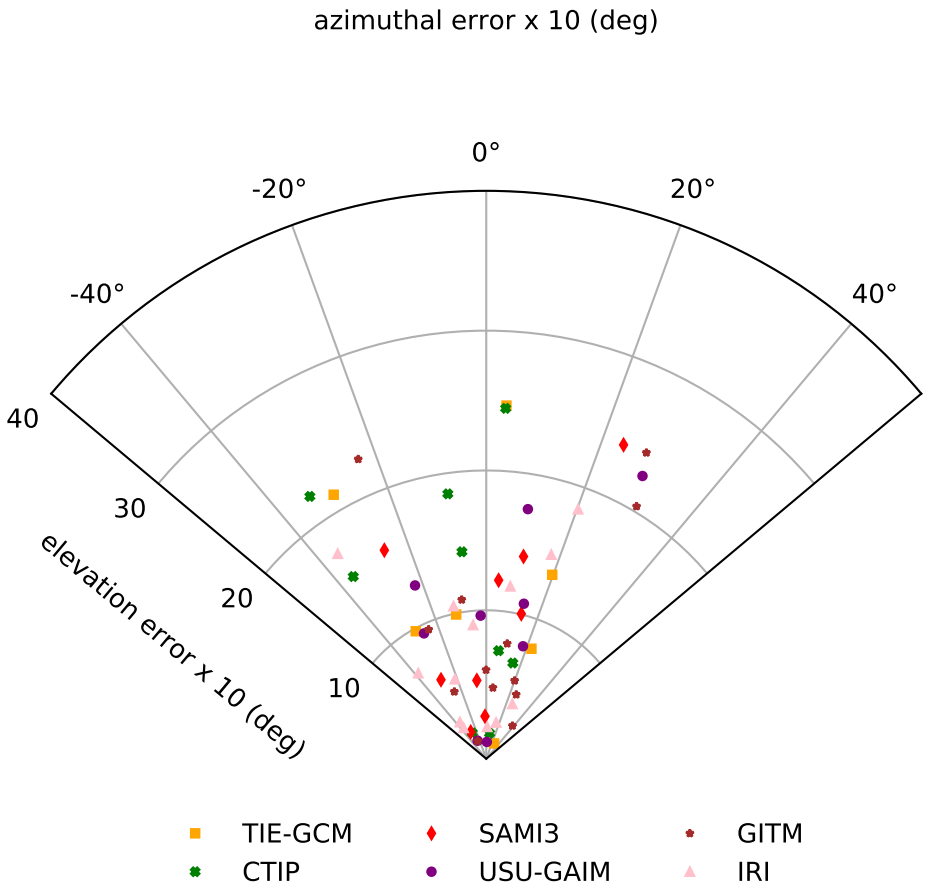

Figure 22. Angle of Arrival error for the POL-G10 link can show a model's preference to where the signal landed. For example, GITM seemed to land eastward of the receiver as most results are seen east of 0°. The errors are × 10 for visual ease.

Figure 22 is a different representation of the AoA error data from before. This plot shows the AoA error data splitting the results into their azimuth and elevation components. We already calculated the error for azimuth and elevation, but we are revealing where the deviation is located in relation to the measured AoA. Each data point is one time for each model. In other words, you will see that are fewer TIE-GCM markers compared to IRI due to the time resolution. In order to interpret the data, you look at the coagulation of data nearest to the 0,0 point (the measured AoA) which marks a perfect hit. It is important to note that given the relatively small distance each beam had to travel latitudinally/longitudinally for this experiment, the errors were near 1°. To display the data so it is readable, we multiplied the errors by 10 so the plot would expand and each point is clearer. Looking deeper into Figure 22, we see that IRI has all of its markers within 2° of error while GITM has a larger spread of values regarding elevation error values though most points favor 2° of azimuth error.

Another way to view the overall path of a model's ray compared to that of observed values is to plot the rays compared to zenith. Inspired by Figure 3, we can display AoA data in another way. Figure 23 is the actual AoA information as opposed to a value indicated by a receiver. It may not depict a quantifiable error like MAE, but for a quick glance at model performance, this guides the eyes to accuracy. The observed values are focused around the location of the transmitter. Looking back at Figure 2, notice that POL is almost directly due North of the transmitter. As a result, the observed beam, as well as the model values, is almost exactly in line with 0°. There is expansion in the elevation values, which are explained by the dynamic nature of the ionosphere over time (Figure 8), but we see that the models which are closer to the observed results seem to also match the EDPs from Figure 8.

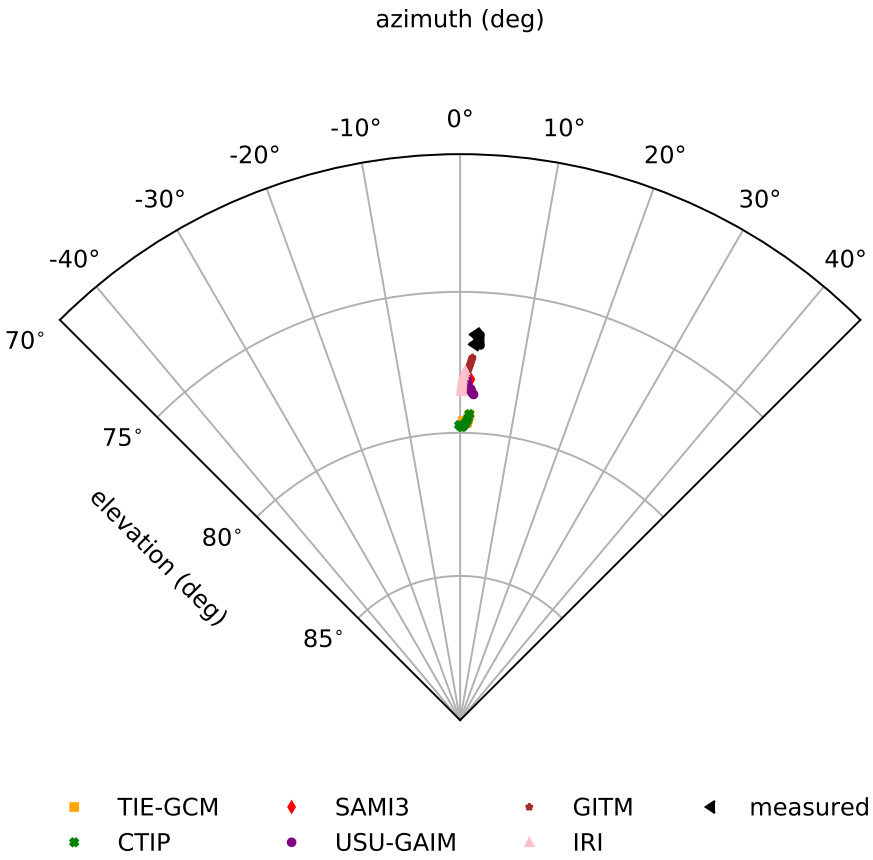

Figure 23. The measured and simulated AoA for the POL-G10 link.

Up to now, we have considered one single transmitter-receiver link. The other 5 links are in a tight enough vicinity around POL-G10 that any deviation seen on the earlier plots is a mostly minuscule amount therefore the decision was made to assume all sites can be treated as equal. With this assumption, we are able to append information and combine the results from all 6 links in an attempt a more statistically sound measurement of performance. This compilation is displayed in Figure 22.

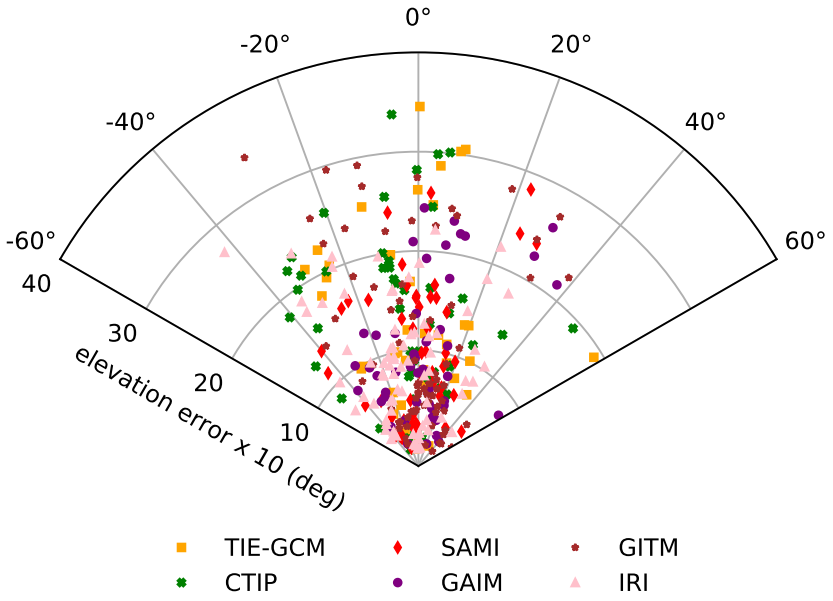

Figure 24. The combined AoA error of all 6 links with errors being multiplied by 10 to allowing for plot clarity. A majority of the points are seen near the 0,0 point indicating that models overall did well.

Once again, we see the errors have been multiplied by 10 in order to widen the results and create a cleaner plot. Interpreting Figure 24 yields similar results with IRI having most of its points near the measured AoA and TIE-GCM having a wider spread in elevation angle.

From this point, we seek to utilize Shim et al. (2011) and the work they did for additional metrics including ratio range and prediction efficiency discussed in Section 2.3. The bolded values in Tables 2, 3, 4, and 5 signify the value which nearest matches the "best" value for that statistic. For the MAE, a MAE of 0 indicates a perfect match to the observations. For Correlation Coefficient, a value of 1 indicates a perfect match between the observations and model values. For Ratio Range, values closest to 1 indicate the model range perfectly matched the range of the observations.

Prediction Efficiency is maximized for those values closest to 0 indicating an efficient model.

Table 2. MAE of all links.

Model	Elevation (deg)	Azimuth (deg)	Group Delay (ms)
TIE-GCM	1.37	**1.56**	0.18
CTIPe	1.37	1.60	**0.05**
SAMI3	1.01	**1.56**	0.07
USU-GAIM	1.00	1.59	0.06
GITM	1.10	1.70	0.13
IRI	**0.97**	1.92	0.09

Table 3. Correlation Coefficient of all links.

Model	Elevation	Azimuth	Group Delay
TIE-GCM	0.770	**0.935**	0.373
CTIPe	**0.838**	0.928	0.522
SAMI3	0.790	0.928	-0.238
USU-GAIM	0.747	0.927	-0.713
GITM	0.793	0.917	**0.676**
IRI	0.796	0.916	0.606

Table 4. The Ratio Range across all links.

Model	Elevation	Azimuth	Group Delay
TIE-GCM	0.096	0.107	0.403
CTIPe	0.136	**0.162**	0.403
SAMI3	0.061	0.110	0.134
USU-GAIM	0.162	0.102	0.336
GITM	0.209	0.079	0.604
IRI	**0.235**	0.056	**0.672**

Table 5. Prediction Efficiency across all links.

Model	Elevation	Azimuth	Group Delay
TIE-GCM	**0.046**	**-1.835**	-2.993
CTIPe	**0.046**	-1.910	**-0.196**
SAMI3	0.297	**-1.835**	-0.575
USU-GAIM	0.304	-1.889	-0.349
GITM	0.231	-2.091	-1.961
IRI	0.326	-2.475	-1.103

From Table 3, USU-GAIM has a strong negative correlation for group delay which matches what was seen in figure 12. SAMI3 also has a negative correlation, but this is a result of stagnation between 19-20Z while the observations steadily climbed. Table 4 shows a cumulative Ratio Range. To calculate this, we took the average of the model values and the average of the observed values separately in order to establish a true average value of the model's range and average value of the observed values.

The equation is then

$$Cumulative\ Ratio\ Range_{(max-min)} = \frac{\langle (x_{mod})_{max} - (x_{mod})_{min} \rangle}{\langle (x_{obs})_{max} - (x_{obs})_{min} \rangle}. \tag{14}$$

Table 5 uses the reference value which is the average value of the observations, therefore we are determining how drastically the observations vary from peak to trough. Alo worth noting, the prediction efficiency we used takes MAE into account as opposed to RMSE like Shim et al. (2011). This equation is then:

$$PE_{MAE} = 1 - \frac{MAE_{mod}}{MAE_{ref}} = 1 - \frac{\Sigma(|x_{mod} - x_{obs}|)/N}{\Sigma(|x_{ref} - x_{obs}|)/N}. \tag{15}$$

3.3 Solid Angle

Throughout these metrics, we have been looking at the three separate parameters of elevation, azimuth and group delay, but clarity is lacking with how to properly interpret performance. The question of if elevation or azimuth accuracy has more weight is not an easy one to answer. As a result, we recast the AoA data to a single spatial parameter, solid angle. To find the solid angle we use the following equations. We start off with the defining arc length, d,:

$$d = \Theta * R \tag{16}$$

where Θ is the angle subtended and R is the radius. Solving for Θ provides:

$$\Theta = d/R = \text{acos}(\sin\phi_1 \times \sin\phi_2 + \cos\phi_1 \times \cos\phi_2 \times \cos\Delta\lambda), \tag{17}$$

where ϕ is the elevation angle , λ is azimuth, the subscript 1 is the model value, and subscript 2 is the measured value. The angle Θ is then used to calculate the solid angle Ω:

$$\Omega = 2\pi(1 - \cos\Theta) \tag{18}$$

Using these equations, we now have a condensed measurement for spatial analysis. With solid angle, we can repeat error, MAE and total MAE for the solid angle measurement.

There is one more graphical representation, discussed in Nickisch et al. (2015), that can be used to determine the feasibility of a model called the Cumulative Distribution Function (CDF). Simply put, CDF shows how much error you would have to accept to envelop a percentage of the total number of data points. In Figures 25 and 26, we show two particular points for reference: one standard deviation and two standard deviations. Looking deeper at Figure 25, if you are only concerned about getting a simple Gaussian majority (1 standard deviation) USU-GAIM reached that value with the least amount of error followed by CTIPe, SAMI3, and IRI with GITM and TIE-GCM requiring more distance to encompass 68%. However, if you seek an overwhelming majority of points (2 standard deviations), the order is changed to CTIPe coming in with the least error followed by USU-GAIM, IRI-2016, then SAMI3 with GITM and TIE-GCM still requiring more error before reaching 95% of points.

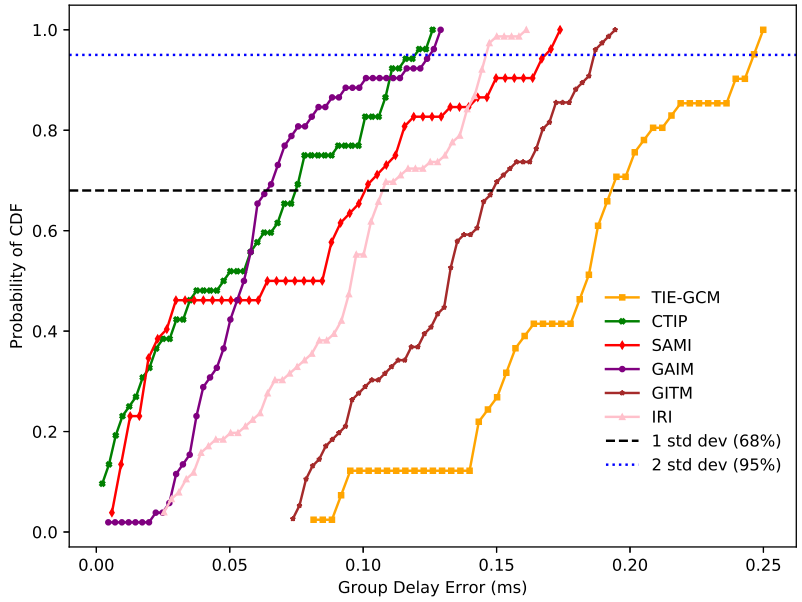

Figure 25. The Group Delay CDF evaluated over all links.

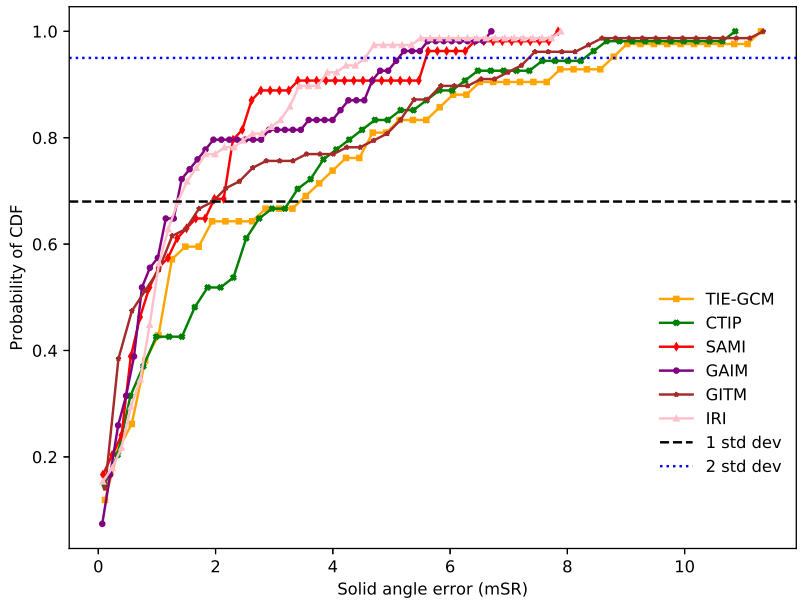

Figure 26. The Solid Angle CDF evaluated including all links

Figure 25 reveals that some models started out closer to the correct value (minimal error) whereas GITM and TIE-GCM started off with more than 0.05 ms of error before you even reached your first points. Figure 26 had some models encompass a majority of their values within 2 mSR.

3.4 X-Mode

To this point, we have determined the accuracy and performance of the models while looking at solely the O-mode, but we are not limited in exploring available data. Going back to Appleton-Hartree equations discussed in Jones and Stephenson (1975), one examined situation is the index of refraction for a cold, collision-less plasma,:

$$n^2 = 1 - \frac{X}{1 - \frac{(Y\sin\psi)^2}{2(1-X)} \pm \sqrt{\frac{(Y\sin\psi)^4}{4(1-X)^2} + (Y\cos\psi)^2}} \tag{19}$$

where X = f_p^2/f^2, Y = f_g/f, and ψ represents the angle between the background magnetic field and the k vector and f is the wave frequency, f_p is the plasma frequency, and f_g is the gyrofrequency. The $\pm$ sign in the denominator is distinguishes between the O-mode (ordinary polarization) and X-mode (extraordinary polarization) with the O-mode positive and the X-mode negative. The X-mode is more susceptible to alterations in the magnetic field as group velocities deflect $\perp$ or $\parallel$ to the magnetic field depending on if f_p^2 is greater than or less than f^2 - f_g^2 respectively (Dao et al., 2016). With most systems capable of obtaining the X-mode data from your ground truth measurements, you will be able to repeat the process, but dealing with X-mode values.

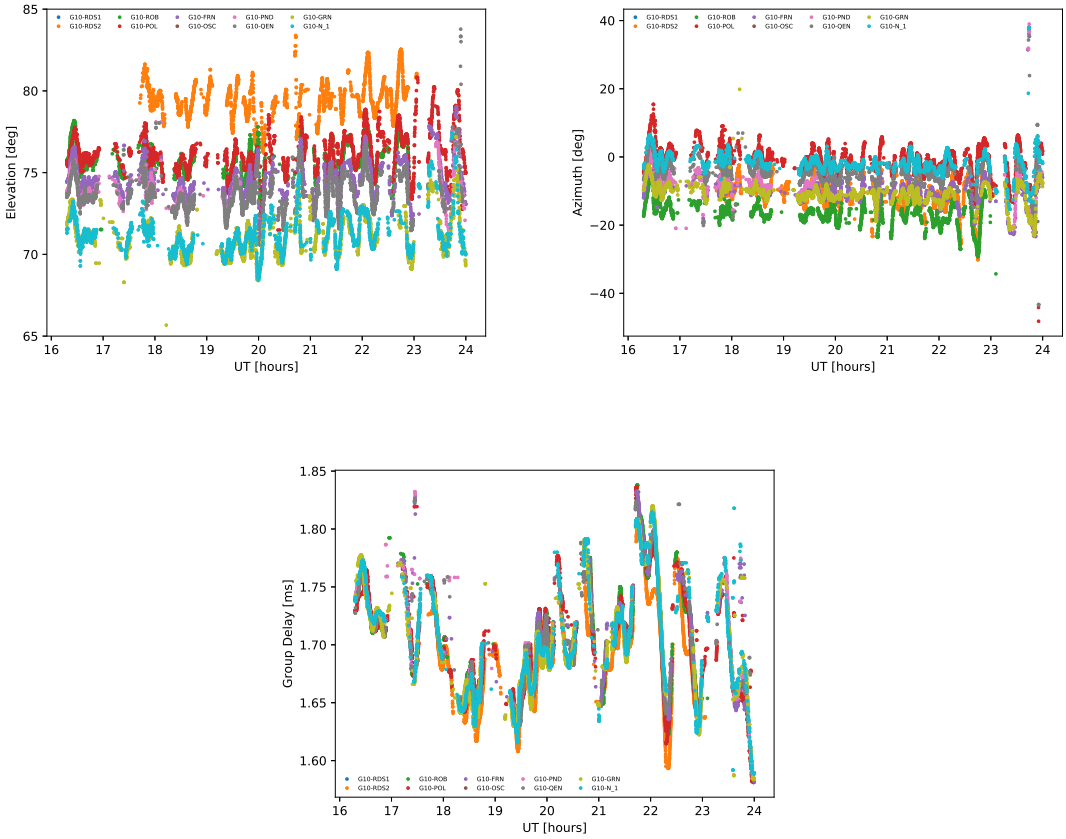

Figure 27. The received X-mode data for elevation(top-left), azimuth (top-right) and group delay (bottom) from WSMR experiment for a 5.33 MHz signal on 22 Jan, 2014.

It is worth mentioning that the EDP for each model doesn't alter because the ionosphere used for CCMC isn't changing. With this in mind, we simply have to refer to Figure 8 for the different interpretations of the EDP for each model. For the sake of comparison, I will be using the same link (POL-G10) as before.

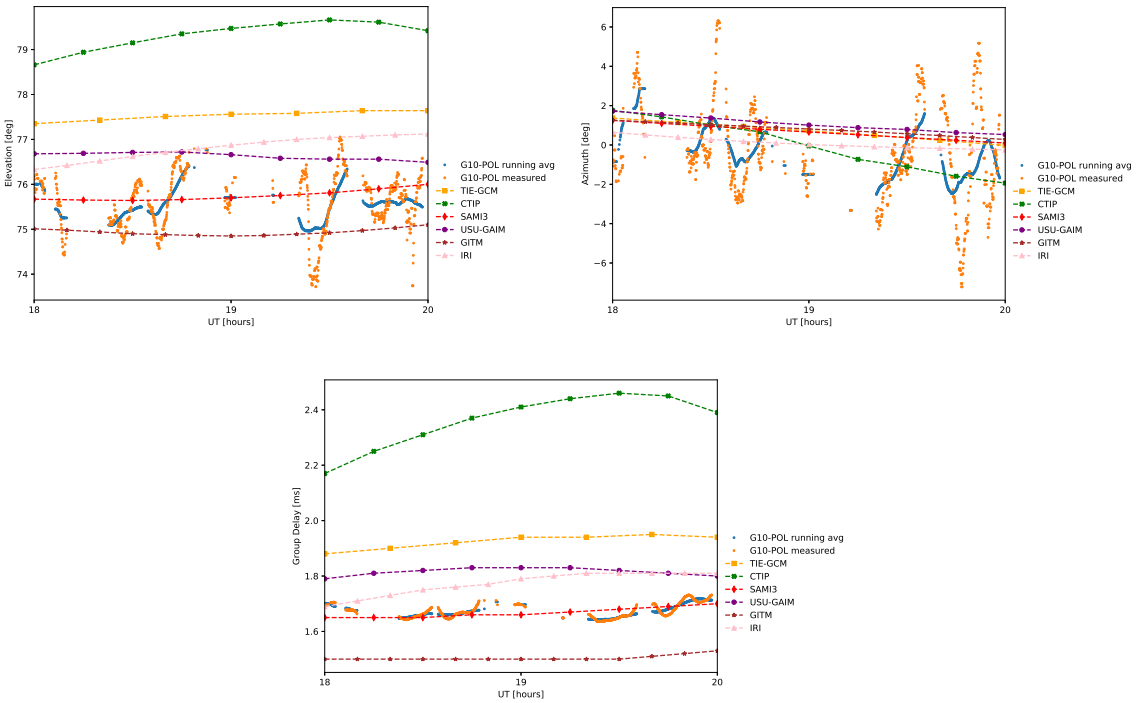

Figure 28. Comparing the X-mode model and observed results for elevation (top-left), azimuth (top-right), and group delay (bottom) for the POL-G10 link on 22 Jan 2014.

X-mode elevation angle data shows a mostly stagnant trend, but it is best mirrored by SAMI3 whereas IRI has a steadily ascending trend. There is a much larger spread between each model when compared to the O-mode in Figure 11. X-mode azimuth data shows a gradual descending trend within both the observations and the models, but CTIPe descends more steeply around 1845Z. Going back to Figure 8, CTIPe had a much more pronounced E-region. This may be the cause for the drastic shift in the received bearing of the signal as there is a more pronounced refractive index in the area. X-mode group delay data shows a wider spread between each model when

compared to Figure 12, but this is exacerbated by CTIPe's enhanced E-region.

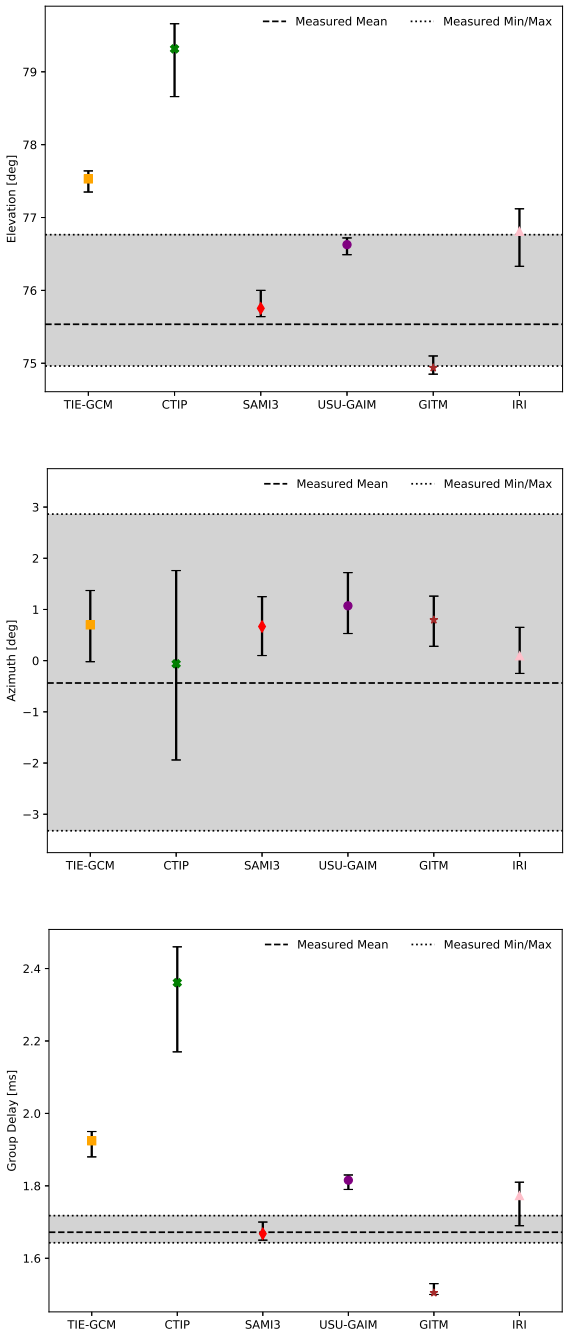

Figure 29. X-mode statistics for elevation (top), azimuth (center), and group delay (bottom)fro the POL-G10 link on 22 Jan 2014.

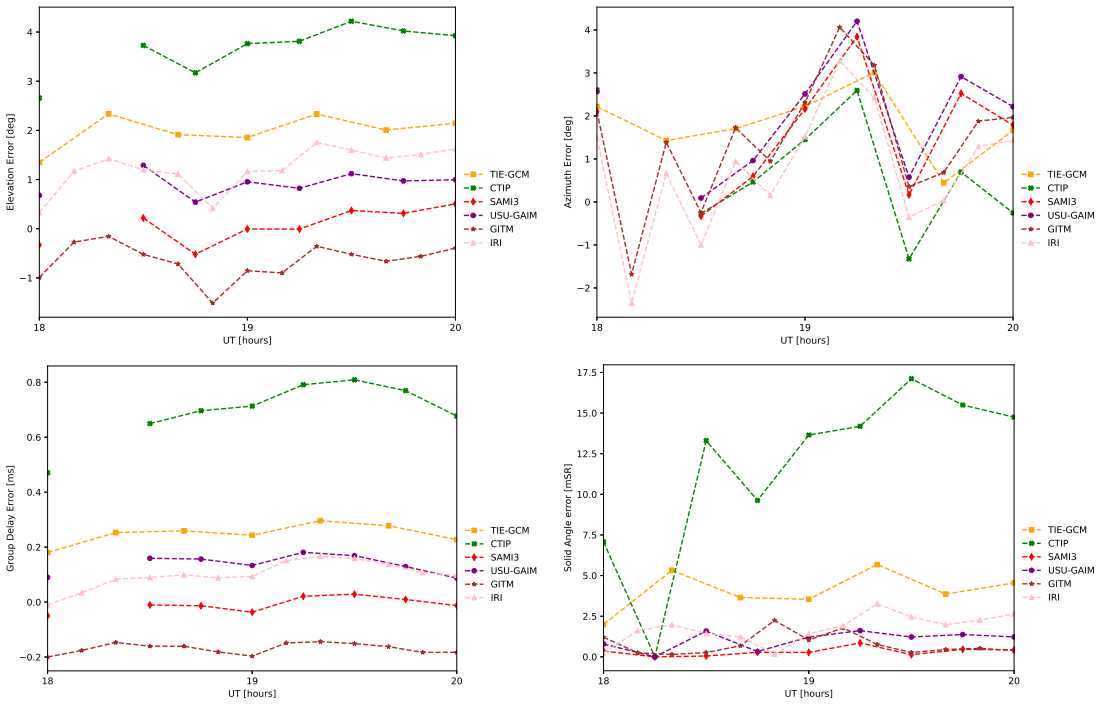

Figure 30. X-mode error over time for elevation (top-left), azimuth (top-right), group delay (bottom-left), and solid angle (bottom-right).

Figure 29 reveals several models which fell within the range of observations for each parameter. Group delay's low-amplitude variance in Figure 28 is matched by several models despite not falling within 1.7 ms. Figure 30 shows an evident absense of data between 18Z and 19Z. The lack of error for CTIPe, USU-GAIM, and SAMI3 is a result of the unavailable data from Figure 28. A possible solution to this issue is to increase the size of the rolling average window and encapsulate more measured data. This will include more points, but may run the risk of introducing too much variability as opposed to minimizing the effect of the outliers.

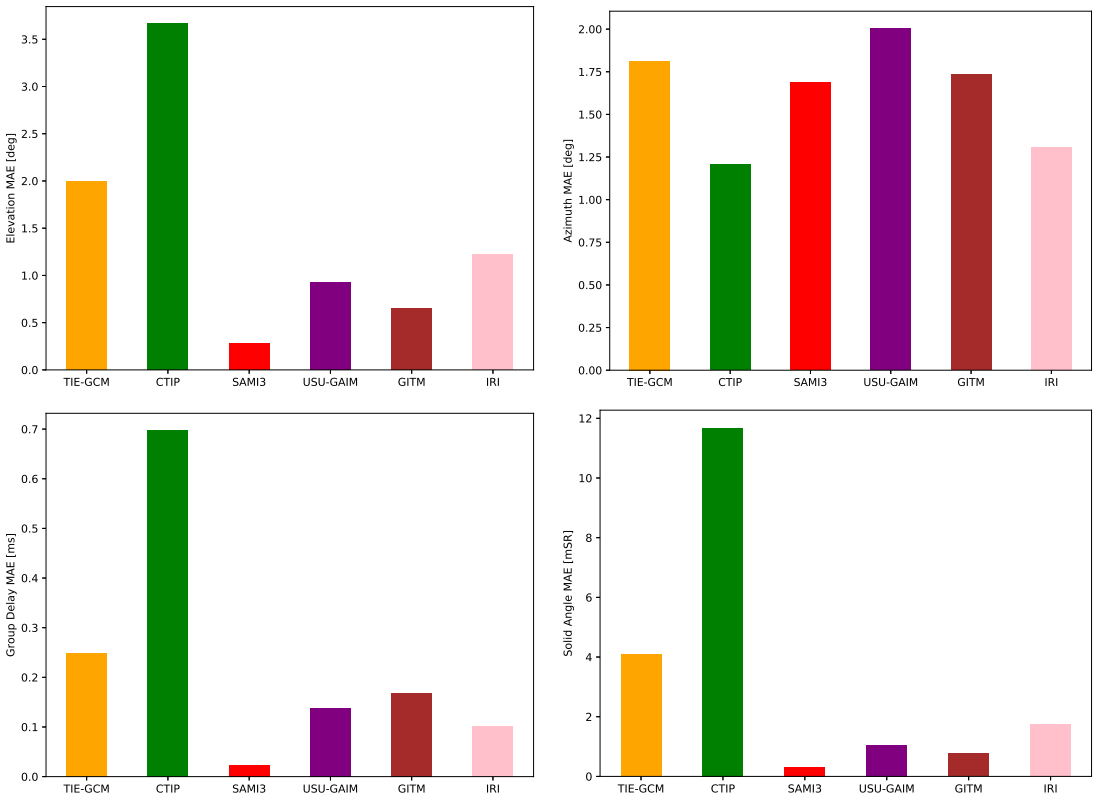

Figure 31. X-mode MAE for elevation (top-left), azimuth (top-right), group delay (bottom-left), and Solid Angle (bottom-right).

Looking back at Figures 19, 20, and 21, we see that SAMI3 performed adequately by comparison for the O-mode, but in regards to the X-mode, SAMI3 performed well. There is also a visual clue that solid angle results appear to more similarly match elevation angle. If you note the performance of the different models on elevation angle, CTIPe has the largest MAE followed by TIE-GCM, IRI, and GAIM. Comparing this order to the solid angle, the results are the exact same. Counter this with azimuth MAE where the order from highest to lowest MAE (GAIM, TIE-GCM, GITM, SAMI3, IRI, and finally CTIPe), it may add some additional insight to the solid angle parameter.

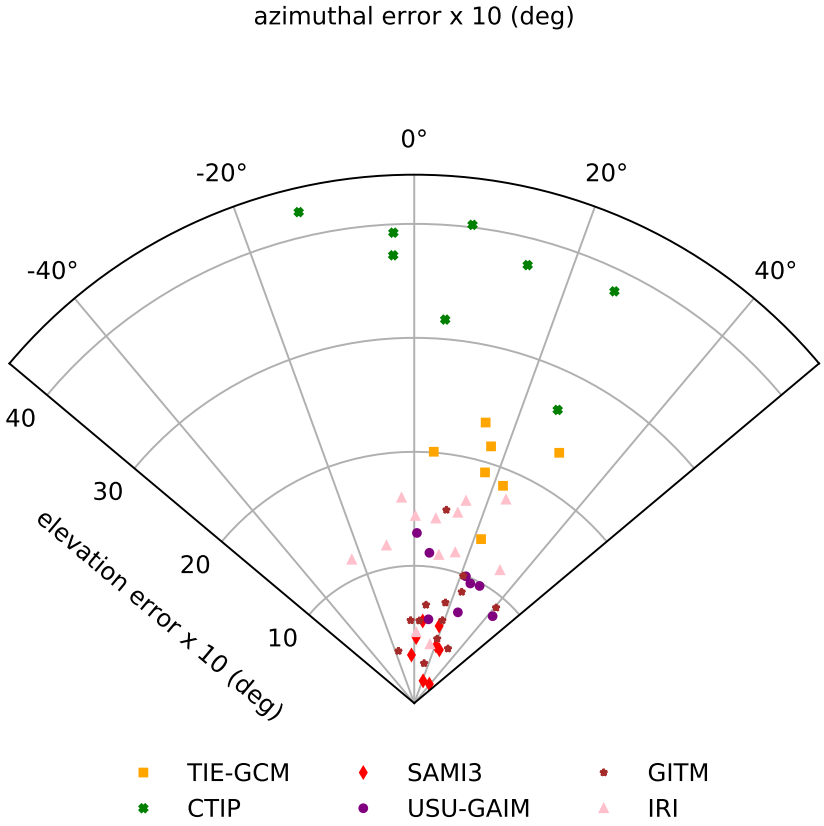

Figure 32. X-mode Radar plot of azimuth and elevation error for POL-G10 link from 22 Jan 2014. Data has been multiplied by 10 for more legible data.

Figure 32 has a couple features of interest. One of the noticeable trends appears to be a more consistent concentration of each model in one general area. In this case, each elevation error value for TIE-GCM is in the same vicinity as other TIE-GCM values. This is repeated for for each model. Also, CTIPe's amplified E-region continues to separate itself from other models. There also appears to be a more consistent concentration of each model in one general area. Another overall trend is the apparent eastward orientation to the received data. Compare this image to Figure 22, where there was a wide spread $\pm$ 3° from North.

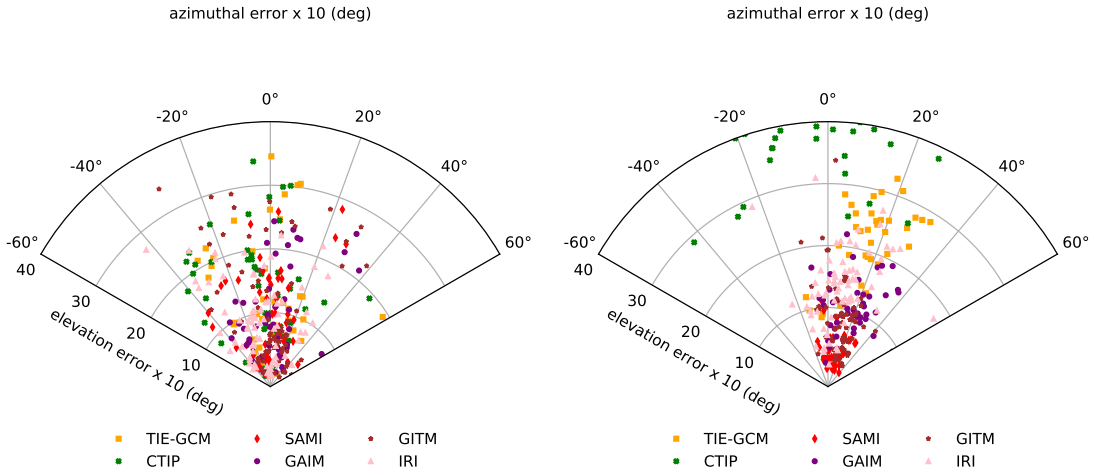

Figure 33. Comparing the cumulative O-mode and X-mode radar plots. The O-mode (left) AoA error has a larger spread between each point when compared to the tight concentration of each model in the X-mode (right).

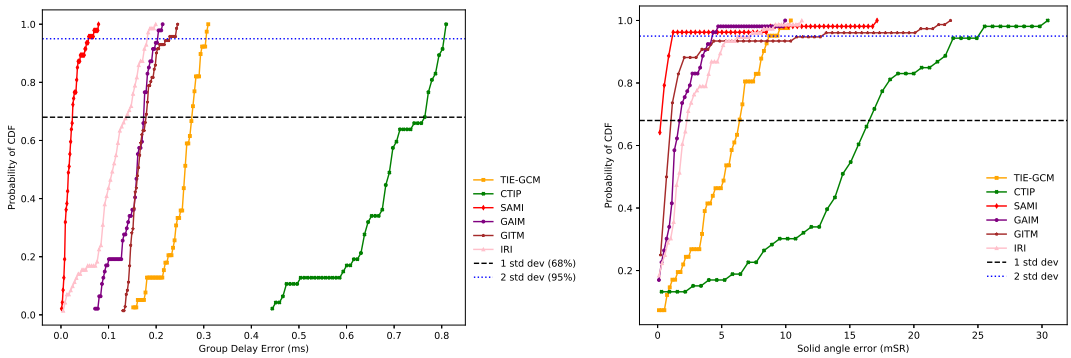

Figure 34. Group Delay CDF (left) and Solid Angle CDF (right) for the X-mode for all links.

Figure 34 shows the CDF for both group delay and solid angle where the remarkable job SAMI3 did with minimizing the error overall all data points is evident. It required less than 0.1 ms of error to incorporate 100% of values and reached 95% of values within 5 mSR. The performances of IRI, GAIM, and GITM were very close to each other for both temporal and spatial measurements. Throughout all these plots,

we've seen SAMI3 handle the X-mode far better than the other models while CTIPe over analyzed the EDP and the magnitude of the error is apparent. It is important to know that these results and plots are for one frequency over one time frame at one location. This is not an indictment or indicator of any one particular model's accuracy.

Tables 6, 7, 8, and 9 reveal the quantifiable metrics indicating performance. Combined, they reveal that SAMI3 had the least error for several parameters, including group delay and solid angle, and had the strongest correlation for elevation. Other models may have performed well in a couple different metrics, but SAMI3 had a consistent success rate over different methods. The bolded values, similar to Tables 2, 3, 4, and 5, reveal the model who most closely matched the "perfect" performance (MAE of 0, Correlation Coefficient of 1, Ratio Range of 1, and Prediction Efficiency of 0).

Table 6. MAE for X-mode over all links.

Model	Elevation (deg)	Azimuth (deg)	Group Delay (ms)	Solid Angle (mSR)
TIE-GCM	2.18	2.32	0.18	4.88
CTIPe	3.88	1.73	0.68	13.52
SAMI3	**0.46**	1.96	**0.02**	**0.81**
USU-GAIM	1.16	2.23	0.15	1.69
GITM	1.10	1.99	0.17	2.11
IRI	1.43	**1.67**	0.11	2.28

Table 7. Values of Correlation Coefficient for X-mode covering all links.

Model	Elevation	Azimuth	Group Delay
TIE-GCM	0.862	0.836	-0.201
CTIPe	0.858	**0.852**	-0.343
SAMI3	**0.915**	0.850	0.114
USU-GAIM	**0.915**	0.850	-0.662
GITM	0.856	0.843	**0.434**
IRI	0.800	0.846	-0.241

Table 8. X-mode results of the Ratio Range covering all links.

Model	Elevation	Azimuth	Group Delay
TIE-GCM	0.116	0.112	0.629
CTIPe	**0.385**	**0.321**	2.935
SAMI3	0.140	0.091	0.524
USU-GAIM	0.133	0.094	0.315
GITM	0.099	0.076	0.210
IRI	0.351	0.076	**1.258**

Table 9. Prediction Efficiency for the X-mode using all links.

Model	Elevation	Azimuth	Group Delay	Solid Angle
TIE-GCM	-0.563	**0.037**	-4.894	0.996
CTIPe	-1.788	0.282	-14.823	**0.988**
SAMI3	0.671	0.187	**0.496**	0.999
USU-GAIM	0.164	0.072	-2.534	0.999
GITM	0.231	0.172	-2.924	0.998
IRI	**-0.025**	0.307	-1.548	0.998

One additional point of error for the models can come from a victim of circumstance. The frequency of interest is 5.33 MHz, but if you reexamine the ionogram (Figure 4), there is an X-mode F_1 cusp indicated on the ionogram near 5.33 MHz. The ray path for signal frequencies near a cusp is extremely sensitive to the shape of the profile leading up to the cusp. If you models can't capture the cusp or shape properly, the simulated rays will show a bias.

IV. Conclusions

A new technique for validating modeled bottomside ionosphere has been presented in this analysis. The true strength of the process is the wide range of metrics and analytical tools. The attempt to minimize the impact of outliers using a running average and mean absolute error allowed the models to have a fair attempt of matching an observed value. Recasting AoA data to one single parameter of solid angle allowed for a quantifiable measurement that simplified the spatial resolution aspect into one value as opposed to two. Qualitative performance can be determined through graphs of AoA Error. Provided the data is not drastically impacted by ionospheric disturbances, this technique can be used to validate performance during calm conditions.

Ultimately, the question for performance relies on what aspect of the model's performance is being analyzed. The statistics have SAMI3 performing very well with average values near that of the observations closely followed in performance by IRI and GAIM. GAIM specifically did well with regards to group delay as it was near a perfect match. TIE-GCM and CTIP, however, were among the models that didn't quite work well with matching the range of the observed values. It is worth noting that CTIP performed better than GITM with respect to group delay. While no model excelled over all metrics, GAIM performed well overall. Though it may not have ranked highest with regards to spatial performance, it was among the best and clearly exceeded the performance of other models when looking at the temporal parameter. This seems to make sense when you take into account that the data assimilation available to GAIM would allow it to adjust results to closer represent the ingested information.

The statistical performance of the model was one measurement of success, but it left out a key detail which was the overall trend of the available information: GAIM

clearly had a descending trend for elevation and group delay while the observations were increasing over time. For this reason, we take a look at the correlation coefficient results where CTIPe captured the elevation trends the best and still represented the azimuth results well. TIE-GCM was able to be the highest performer for the azimuth data, but did not show a strong correlation for the group delay information which is one category GITM did do well in. IRI is among the top performers across the board as it is top 3 in elevation and a near perfect azimuth correlation.

Error is a simple measurement for success whereas the model who deviates the least from the obs represents the most accurate choice. Looking at the MAE, GAIM is among the best performers consistently though never directly being reported as the most accurate which goes to IRI. The solid angle MAE (mSR) for the models were: TIE-GCM: 2.62, CTIPe: 2.58, SAMI3: 1.60, GAIM: 1.55, GITM: 2.14, IRI: **1.50**. These seem to indicate that the elevation angle is a stronger driving force or indicator of how well a model would perform with regards to the solid angle as IRI is once again the most accurate model followed closely by GAIM and SAMI3.

If accuracy of the model isn't a major concern, but were more interested in how well a model represents the variation in observations, utilization of the metrics ratio range and prediction efficiency are ideal. From here, we see the climotological model IRI matches the observations with respect to both elevation and group delay followed by GITM. This not the case for prediction efficiency as CTIPe rates the highest in elevation and group delay. TIE-GCM did well in elevation while performing the highest in azimuth. Once again, the large discrepancy in its EDP could explain where difference in performance between spatial and temporal parameters may come from.

If the main concern is the magnitude of error one is willing to accept to obtain a majority of data, than the CDF is the go-to metric. In the group delay CDF, GAIM may have been the first model with a simple majority as it only requires 0.05 ms error

to encapsulate 68% of observations, but CTIPe overtook GAIM before reaching 95% of observations. For solid angle, IRI required the least error to reach both standard deviations, but it was tightly contested with GAIM and SAMI3.

The many measurements, plots, and tables reveal that no one model is superior to all others across all methods and therein lies the strength of this technique. You require all information and results before rushing towards a hasty decision. Also, looking solely at the "top" performer may restrict your focus away from a consistent performer which is comparable to always winning the silver medal as opposed to one gold and not placing in other events. For this reason, over this data set, the most accurate models would be USU-GAIM followed by IRI. It is important to emphasize this thesis introduces a technique and nothing more. The decision for rankings came with a greater weight of importance to evaluating the O-mode rather than equal O-mode/X-mode importance. If you adjust the weighting equally, SAMI3's capabilities were the most consistent across the different polarizations. It is vital to reiterate, the results discussed in this thesis do not represent the overwhelming abilities of the models provided. They are a snapshot of one 2 hour period of one day in one location while looking at one frequency.

For a true validation effort, a larger data set is required with regards to testing over solar maximum/solar minimum, different seasons within those solar maximums/minimums, different latitudes/longitudes to compare performance of low/middle/high latitude runs as well as seeing what effect magnetic longitude may have on some models. In addition to altering the temporal length of an experiment, one can adjust the spatial scale through the use of an established transmitter and setting up numerous receivers capable of measuring AoA data a great distance from said transmitter. A large geographic spread adds testing the global capabilities to the models as opposed to the minor distance detailed in this experiment.

Bibliography

Anderson, D. N., Vuonsanto, M. J., Codrescu, M., Decker, D., Fesen, C. G., Fuller-Rowell, T. J., Reinisch, B. W., Richards, P. G., Roble, R. G., Schunk, R. W., and Sojka, J. J. (1998). Intercomparison of Physical Models and Observations of the Ionosphere. *Journal of Geophysical Research*, 103(A2):2179–2192.

Araujo-Pradere, E. A. and Fuller-Rowell, T. J. (2002). STORM: An Empirical Storm-time Ionospheric Correction Model. *Radio Science*, 37(5).

Bilitza, D., Altadill, D., Truhlik, V., Shubin, V., Galkin, I., Reinisch, B., and Huang, X. (2007). International Reference Ionosphere 2016: From Ionospheric Climate to Real-time Weather Predictions. *Space Weather*, 15:418–429.

Chulaki, A. (2019). CCMC Hosted Models at a Glance. `https://ccmc.gsfc.nasas.gov/models/models_at_glance.php`.

Codrescu, M. V., Fuller-Rowell, T. J., Munteanu, V., Minter, C. F., and Millward, G. H. (2008). Validation of the Coupled Thermosphere Ionosphere Plasmasphere Electrodynamics model: CTIPE-Mass Spectrometer Incoherent Scatter Temperature Comparison. *Space Weather*, 6.

Dao, E. V., McNamara, L. F., and Colman, J. J. (2016). Magnetic Field Effects on the Accuracy of Ionospheric Mirror Models for Geolocation. *Radio Science*, 51:284–300.

Decker, D. T. and McNamara, L. F. (2007). Validation of Ionospheric Weather Predicted by Global Assimilation of Ionospheric Measurements (GAIM) models. *Radio Science*, 42.

Ferguson, J. A. (1995). Ionospheric Model Validation at VLF and LF. *Radio Science*, 30(3):775–782.

Garcia-Sage, K. (2019a). Coupled Thermosphere Ionosphere Plasmasphere Electrodynamics Model. https://ccmc.gsfc.nasas.gov/models/modelinfo.php?model=CTIPe.

Garcia-Sage, K. (2019b). Global Ionosphere Thermosphere Model (GITM). https://ccmc.gsfc.nasas.gov/models/modelinfo.php?model=GITM.

Garcia-Sage, K. (2019c). International Reference Ionosphere Model - IRI (2016). https://ccmc.gsfc.nasas.gov/models/modelinfo.php?model=IRI.

Garcia-Sage, K. (2019d). SAMI3 Low-to-Mid-Latitude Ionosphere. https://ccmc.gsfc.nasas.gov/models/modelinfo.php?model=SAMI3.

Garcia-Sage, K. (2019e). Thermosphere Ionosphere Electrodynamics General Circulation Model. https://ccmc.gsfc.nasas.gov/models/modelinfo.php?model=TIE-GCM.

Garcia-Sage, K. (2019f). Utah State University Global Assimilation of Ionospheric Measurements (USU-GAIM) Model. https://ccmc.gsfc.nasas.gov/models/modelinfo.php?model=GAIM.

Huba, J. D., Joyce, G., and Krall, J. (2008). Three-dimensional Equatorial Spread F Modeling. *Geophysical Research Letters*, 35.

Jones, M. R. and Stephenson, J. J. (1975). A Versatile Three-dimensional Ray Tracing Computer Program for Radio Waves in the Ionosphere OT Rep. 75-76U. *S. Dep. of Commer., Office of Telecommun.*

McNamara, L. (1985). The Use of Total Electron Content Measurements to Validate Empirical Models of the Ionosphere. *Advances in Space Research*, 5:81–90.

McNamara, L. F. (1991). *The Ionosphere: Communications, Surveillance, and Direction Finding.* Krieger publishing company.

Nickisch, L. J., Fridman, S., Hausman, M., Kraut, S., and Zunich, G. (2015). Assimilative Modeling of Ionospheric Dynamics for Nowcasting of HF Propagation Channels in the Presence of TIDs. *Radio Science*, 51:184–193.

Reinisch, B. W. & Galkin, I. A. (2011). Global Ionospheric Radio Observatory (GIRO). *Earth, Planets and Space*, 63(4):377–381.

Richmond, A. D., Ridley, E. C., and Roble, R. G. (1992). A Thermosphere/Ionosphere General Circulation Model with Coupled Electrodynamics. *Geophysical Research Letters*, 6(6):601–604.

Ridley, A. J., Deng, Y., and Toth, G. (2006). The Global Ionosphere-Thermosphere Model. *Journal of Atmospheric and Solar-Terrestrial Physics*, 68(8):839–864.

Scherliess, L., Schunk, R. W., Sojka, J. J., Thompson, D. C., and Zhu, L. (2006). Utah State University Global Assimilation of Ionospheric Measurements Gauss-Markov Kalman Filter Model of the Ionosphere. *Journal of Geophysical Research*, 111.

Schunk, R. W., Scherliess, L., Sojka, J. J., Thompson, D. C., Anderson, D. N., Codrescu, M., Minter, C., Fuller-Rowell, T. J., Heelis, R. A., Hairston, M., and Howe, B. M. (2004). Global Assimilation of Ionospheric Measurements (GAIM). *Radio Science*, 39.

Shim, J. S., Kuznetsova, M., Rastatter, L., Hesse, M., Bilitza, D., Butala, M., Codrescu, M., Emery, B., Foster, B., Fuller-Rowell, T., Huba, J., Mannucci, A. J.,

Pi, X., Ridley, A., Scherliess, L. Schunk, R. W., Stephens, P., Thompson, D. C., Zhu, L., Anderson, D., Chau, J. L., Sojka, J. J., and Rideout, B. (2011). CEDAR Electrodynamics Thermosphere Ionosphere (ETI) Challenge for Systematic Assessment of Ionosphere/Thermosphere Models: NmF2, hmF2, and Vertical Drift using Ground-based Observations. *Space Weather*, 9.

Solomon, S. C. (2006). Numerical Models of the E-region Ionosphere. *Advances in Science Research*, 37(5):1031–1037.

Yang, G., Sun, Y., Bai, W., and et al (2018). Validation Results of NmF2 and hmF2 Derived from Ionospheric Density Profiles of GNOS of FY-3C Satellite. *Science China Technological Sciences*, 61(9):1372–1383.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704–0188

1. REPORT DATE *(DD–MM–YYYY)*	2. REPORT TYPE	3. DATES COVERED *(From — To)*
26–03–2020	Master's Thesis	May 2018 — Mar 2020

4. TITLE AND SUBTITLE

VALIDATION TECHNIQUE FOR MODELED BOTTOMSIDE IONOSPHERES VIA RAY TRACING

5a. CONTRACT NUMBER

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)

Burg, Kevin Scott, 1st Lt, USAF

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Air Force Institute of Technology
Graduate School of Engineering and Management (AFIT/EN)
2950 Hobson Way
WPAFB OH 45433-7765

8. PERFORMING ORGANIZATION REPORT NUMBER

AFIT-ENP-MS-20-M-085

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)

Air Force Research Laboratory/Space Vehicles
3550 Aberdeen Avenue SE
Kirtland AFB, NM 87117
DSN 263-2936, COMM 505-846-3172
Email: eugene.dao@us.af.mil

10. SPONSOR/MONITOR'S ACRONYM(S)

AFRL/RV

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT

DISTRIBUTION STATEMENT A:
APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED.

13. SUPPLEMENTARY NOTES

14. ABSTRACT

A new method for validating ionosphere models using High Frequency (HF) angle of arrival (AoA) data is presented. AoA measurements from a field campaign held at White Sands Missile Range, New Mexico, USA in January 2014 provide the actual elevation angle, azimuth and group delay results from 10 transmitter-receiver circuits. Simulated AoAs are calculated by ray tracing through the electron density profiles predicted from the ionosphere models hosted by NASA's Community Coordinated Modeling Center: IRI-2016, USU-GAIM, GITM, CTIPe, TIE-GCM, and SAMI3. Through the implementation of metrics including Mean Absolute Error, Prediction Efficiency, Correlation Coefficient, and others, we are able to compare and analyze model performance using this new model validation approach. The wide variety of metrics determine model performance quantitatively and qualitatively by analyzing behavior over time.

15. SUBJECT TERMS

Model Validation, Bottomside Ionosphere

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			Maj Daniel J. Emmons, AFIT/ENP
U	U	U	U	72	19b. TELEPHONE NUMBER *(include area code)* (937) 255-3636x4571; Daniel.Emmons@afit.edu